Dynamic Modeling, Optimization, and Advanced Control for Large Scale Biorefineries

Prunescu, Remus Mihail

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Technical University of Denmark

Remus Mihail Prunescu

Dynamic Modeling, Optimization, and Advanced Control for Large Scale Biorefineries

PhD Thesis, November 2015

DTU Electrical Engineering
Department of Electrical Engineering

Dynamic Modeling, Optimization, and Advanced Control for Large Scale Biorefineries

Remus Mihail Prunescu

Technical University of Denmark
Kgs. Lyngby, Denmark, 2015

Technical University of Denmark
Automation and Control (AUT)
Elektrovej Building 326
DK-2800, Kgs. Lyngby
Denmark
Phone: (+45) 45 25 35 76
Email: info@elektro.dtu.dk
www.elektro.dtu.dk

ISBN: N/A

Summary

Second generation biorefineries transform agricultural wastes into biochemicals with higher added value, e.g. bioethanol, which is thought to become a primary component in liquid fuels [1]. Extensive endeavors have been conducted to make the production process feasible on a large scale, and recently several commercial size biorefineries became operational: Beta Renewables (Italy, 2014), Abengoa Bioenergy (USA, 2014), POET-DSM (USA, 2014), GranBio (Brazil, 2014) [2], while others are under construction, e.g. the Måbjerg Energy Consortium in Denmark.

This thesis presents the findings of a 3 years PhD project that was run by Technical University of Denmark (DTU) in collaboration with the largest Danish energy company DONG Energy A/S between 2012 and 2015. The company owns a demonstration scale second generation biorefinery in Kalundborg, Denmark, also known as the Inbicon demonstration plant [3]. The goal of the project is to utilize real-time data extracted from the large scale facility to formulate and validate first principle dynamic models of the plant. These models are then further exploited to derive model-based tools for process optimization, advanced control and real-time monitoring.

The Inbicon biorefinery converts wheat straw into bioethanol utilizing steam, enzymes, and genetically modified yeast. The biomass is first pretreated in a steam pressurized and continuous thermal reactor where lignin is relocated, and hemicellulose partially hydrolyzed such that cellulose becomes more accessible to enzymes. The biorefinery is integrated with a nearby power plant following the Integrated Biomass Utilization System (IBUS) principle for reducing steam costs [4]. During the pretreatment, by-products are also created such as organic acids, furfural, and pseudo-lignin, which act as inhibitors in downstream processes. The pretreated fibers consist of cellulose and xylan, which are then liquefied in the enzymatic hydrolysis process with the help of enzymes. High glucose and xylose yields are thus obtained for co-fermentation. Ethanol is recovered in distillation columns followed by molecular sieves for achieving a high concentration ethanol. Lignin is separated in the first column and recovered as bio-pellets in an evaporation unit. The bio-pellets are then burnt in the nearby power plant for steam generation.

The first part of this research presents a large scale dynamic model of the plant, separated in modules for pretreatment, enzymatic hydrolysis, and fermentation. The pretreatment and enzymatic hydrolysis models have been validated and analyzed in this study together with a comprehensive sensitivity and uncertainty analysis [5, 6]. The models embed mass and energy balances with a complex conversion route. Computational fluid dynamics is used to model transport phenomena in large reactors capturing tank profiles, and delays due to plug flows. This work publishes for the first time demonstration scale real data for validation showing that the model library is suitable for optimization, control and monitoring purposes. As an application, the pretreatment dynamic model is used to construct a real-time observer that acts both as a measurement filter, and soft sensor for biomass components that are not measured, e.g. pretreatment inhibitors [5].

The next part of this study deals with building a plantwide model-based optimization layer, which searches for optimal values regarding the pretreatment temperature, enzyme dosage in liquefaction, and yeast seed in fermentation such that profit is maximized [7]. When biomass is pretreated, by-products are also created that affect the downstream processes acting as inhibitors in enzymatic hydrolysis and fermentation. Therefore, the biorefinery is treated in an integrated manner capturing the trade-offs between the conversion steps. Sensitivity and uncertainty analysis is also performed in order to identify the modeling bottlenecks and which feedstock components need to be determined for an accurate prediction. This analysis is achieved with Monte Carlo simulations and Latin Hypercube Sampling (LHS) on feedstock composition and kinetic parameters following the methodology from [5, 6, 8, 9].

In the last part of this work, two applications of the L1 adaptive output feedback controller [10] are developed: one for biomass pretreatment temperature [11] and another one for pH in enzymatic hydrolysis [12]. Biomass conversion is highly sensitive to these process parameters, which exhibit nonlinear behavior and can change nominal values. The adaptive controllers are found to perform better across multiple operational points without the need of retuning.

Resumé

Anden-generations bioraffinaderier omdanner affaldsprodukter fra landbruget til kemiske produkter med højere værdi som f.eks. bioethanol, der i fremtiden forventes at blive en primær komponent i flydende brændsler [1]. Der er sket store fremskridt for at skalere denne produktion og der er i de senere år blevet idriftsat flere kommercielle anlæg: Beta Renewables (Italy, 2014), Abengoa Bioenergy (USA, 2014), POET-DSM (USA, 2014), GranBio (Brazil, 2014) [2], mens andre er under planlægning: herunder et anlæg ved Måbjerg Energy Center.

Denne afhandling præsenterer et 3-årigt PhD projekt som er udført i samarbejde mellem Technical University of Denmark (DTU) og DONG Energy A/S i perioden 2012 til 2015. DONG Energy ejer demonstrationsanlægget Inbicon, som er anden-generation bioraffinaderi i Kalundborg [3]. Projektets formål er bruge anlægsdata fra dette demonstrationsskala anlæg til at beskrive og validere dynamiske proces- og kinetikmodeller af anlægget. Disse modeller bliver så brugt til at udvikle model-baseret værktøjer for procesoptimering, avanceret regulering og direkte overvågning af processen.

Bioraffinaderiet omdanner halm til bioethanol ved brug af damp, enzymer og genmodificeret gær. Halmen bliver forbehandlet i en kontinuert reaktor under højt damptryk, hvor ligninen bliver åbnet og hemicellulosen bliver delvist hydrolyseret. Bioraffinaderiet er integreret med det nærliggende kraftværk for at reducere dampomkostningerne [4]. Under forbehandlingen dannes der bi-produkter som organiske syrer, furfural og pseudo-lignin, som alle er inhibitorer i de efterfølgende processer. De forbehandlede fibre består af cellulose og xylan, som bliver enzymatisk hydrolyseret i dette næste trin til glucose og xylose. Det tredje trin er fermentering, hvor sukkerstofferne omdannes til ethanol. Bioethanolen bliver separeret efter fermenteringen i en distillationskolonne samt molekylesi for at opnå høje ethanol koncentrationer. Ligninen udskilles desuden i den første kolonne og omdannes til pilleform vha. fordamperenhed. Disse ligninpiller kan så forbrændes i kraftværket for yderligere dampproduktion. Den første del af denne forskning præsenterer en stor-skala dynamisk model af anlægget, opdelt i følgende moduller: forbehandling, enzymatisk hydrolyse og fermentering. Modellerne for forbehandlingen og den

enzymatiske hydrolyse er blevet valideret og analyseret sammen med en omfattende sensitivitet og usikkerhedsanalyse [5, 6]. Modellerne inkluderer masse- og energibalancer, samt en kompleks kinetikbeskrivelse af de kemiske reaktioner. Dynamiske strømningsberegninger bruges til at modellere de forskellige transport fænomener internt i reaktorerne. Dette projekt viser for første gang validerede data fra et demonstrationsanlæg, hvor et omfattende modelbibliotek som kan bruges til optimerings-, regulerings- og overvågningsformål. Den dynamiske model for forbehandlingen bruges både som et valideringsværktøj for målinger samt at danne indirekte værdier for vigtige biomasse komponenter som ikke måles under processen, f.eks. inhibitorer [5].

Anden del af projektet omhandler en prisoptimeringsmodel for hele anlægget, som kan optimere for forbehandlingstemperatur, enzymdosering under hydrolyse og gær tilsætning ved fermenteringen [7]. Under forbehandlingen af biomassen bliver der dannet bi-produkter som kan inhibere både i den enzymatiske hydrolyse og fermenteringen. Derfor bliver bioraffinaderiet modelleret samlet, så man kan relatere påvirkninger mellem de enkelte omdannelsesprocesser. Følsomheds- og usikkerhedsanalyse er også udført for at identificere de kritiske modelparametre og hvilke biomassekomponenter som er vigtige for at opnå høj nøjagtighed. Monte Carlo simuleringer og Latin Hypercube Sampling (LHS) er udført for biomassesammensætningen og kinetik parametre i metodikken beskrevet i [5, 6, 8, 9].

I den sidste del af projektet er der udviklet 2 reguleringer af L1 adaptive output feedback controller [10]: en for forbehandlingstemperatur af biomasse [11] og en anden for pH-styring under den enzymatiske hydrolyse [12]. Omdannelsen af biomasse har en stærk afhængighed af disse ikke-lineære parametre som desuden ændrer nominelle værdier. De adaptive reguleringer viser sig at kunne performe bedre over et større driftområde uden brug for rekalibrering.

Preface

This project was prepared as a collaboration between academia and industry within the Industrial PhD program set by the Danish Innovation Fund. The university partners consists of the Department of Electrical Engineering, Automation and Control Group, Technical University of Denmark (DTU), and the Department of Chemical and Biochemical Engineering, CAPEC-PROCESS Group, DTU. The industrial sponsor is the largest Danish energy company, i.e. DONG Energy A/S.

The project advisors were:

- Professor Mogens Blanke (main supervisor), Department of Electrical Engineering, Automation and Control Group, DTU;
- Associate Professor Gürkan Sin (co-supervisor), Department of Chemical and Biochemical Engineering, CAPEC-PROCESS Group, DTU;
- Jakob Munch Jensen (company supervisor 2012-2014), Department of Process Control and Optimization, DONG Energy A/S.
- Jon Geest Jakobsen (company supervisor 2014-2015), Department of Process Control and Optimization, DONG Energy A/S.

The thesis consists of a summary report of all findings, and a collection of published articles in peer reviewed scientific journals and conference proceedings in the period 2012-2015.

Kongens Lyngby, November 2015
Remus Mihail Prunescu

Acknowledgments

I was first introduced to biomass refining in 2011 when I came in contact with the Inbicon technology. Back then I was finalizing my master's studies and I was looking for a final project idea. My university supervisor, Professor Mogens Blanke, presented me to Dr. Tommy Mølbak and Dr. Jakob Munch Jensen from DONG Energy A/S. Together we created a 6 months project that dealt with modeling and control of the thermal reactor in biomass pretreatment. Everything went fine and we decided to continue our collaboration with an Industrial PhD project on a more extended topic that included the entire facility. I take the chance here to thank the industrial partner for their interest into research, and for giving me the opportunity to further develop their technology.

The PhD project has been supervised by four very skillful and dedicated people: Professor Mogens Blanke, Associate Professor Gürkan Sin, Dr. Jakob Munch Jensen and Dr. Jon Geest Jakobsen. I had pursued most of my studies within the Automation and Control Group from the Electrical Engineering Department (AUT) where Professor Mogens Blanke was also affiliated. I met Gürkan Sin at the Chemical and Biochemical Engineering department during one of the university master's courses. It seemed natural to collaborate with Gürkan Sin on both the master and PhD project due to his expertise on biorefinery technology and computer aided process engineering thus creating a fully interdisciplinary work.

Jakob Munch Jensen and Jon Geest Jakobsen have been the company co-supervisors and skillfully showed me how to combine academia with the industry. Working together with Professors Mogens Blanke and Gürkan Sin has been inspiring and productive. I learned a lot from their vast experience and constructive criticism being able to produce high quality results in the end. I am really grateful for the guidance and supervision I received from all my supervisors.

Throughout my employment at DONG Energy A/S I also met very dedicated people whom I'd like to thank for their training and challenges we solved together. Besides Tommy Mølbak, Jakob Munch Jensen, and Jon Geest Jakobsen, I'd like to mention Flemming Mathiesen, Jesper Dohrup, Michael Eleskov, Pia Jørgensen, Kit K. Mogensen, Kristian Livijn and Ningling Rao.

I had a short academic stay at EPFL where I collaborated on process optimization with Professor Dominique Bonvin and Scientist Timm Faulwasser. I'd like to thank them for their interest and for the very productive and efficient research visit.

I'd like thank my colleagues from the Automation and Control Group. I made great friends among them and I'm grateful for all the coffee breaks and other off-work activities we had together.

I've been continuously supported by my family throughout all these years and they believed in me for achieving this task. For that I wish to thank them all, my parents, grandparents, sister, aunts and uncle.

Table of Contents

List of Abbreviations

CFD Computational Fluid Dynamics. 23, 26, 32

CSTR Continuous Stirred Tank Reactor. 28, 32

CSTRs Continuous Stirred Tank Reactors. 17

DTU Technical University of Denmark. i, iii, v, 7

EPFL École Polytechnique Fédérale de Lausanne. 63

GMO Genetically Modified Organisms. 3, 7, 14, 18, 29, 42, 69

HPLC High-Performance Liquid Chromatography. 6, 20, 35, 37, 40

IAE Integral Absolute Error. 10, 56, 69

IAPWS IF97 The International Association for the Properties of Water and Steam - Industrial Formulation 1997. 22

IBUS Integrated Biomass Utilization System. i, 2, 14

LHS Latin Hypercube Sampling. ii, iv, 21, 36, 37, 39, 47

MRAC Model Reference Adaptive Controller. 55

NCO Necessary Conditions for Optimality. 67, 70

NIR Near Infrared. 6, 13, 20, 34, 37, 39, 40, 52, 59

OCP Optimal Control Problem. 11, 64, 65, 67

ODEs Ordinary Differential Equations. 19, 21

PI Proportional Integral. 63

RTO Real-Time Optimization. 67, 68

SISO Single Input Single Output. 56, 60

UDS Upwind Difference Scheme. 24

Chapter 1

Introduction

1.1 Background

Petroleum supplies most of the liquid fuel demand in transportation and industry sectors [13]. Projections show that oil prices will rise in the near future because of an increase in energy consumption as the world continues to develop, and also due to depletion of easily accessible resources. The new reserves require a more advanced and expensive technology to extract, leading to a higher price. Nowadays society depends on oil, which is a limited resource with a long life cycle that will eventually vanish. World oil depletion models show that reservoirs would be exhausted by mid century [14, 15].

Burning fossil fuels along with other industrialized activities such as cattle ranching and deforestation, contributes significantly to the emission of gases with greenhouse effects, which are responsible for global warming. It is nearly impossible to keep the worldwide average temperature increase below 2 °C above pre-industrial times by the end of the century [16, 17]. Studies show that concentration of carbon dioxide in atmosphere has been constantly rising since the industrial revolution with negative effects on climate, such as ice sheets melting, ocean acidification, permafrost melting [18].

In order to counteract the dangers of fossil fuels, and to create a sustainable society, governments and global organizations channeled extensive endeavors into the development of renewable and alternative sources of energy. All these efforts are supported by the Kyoto Protocol, an international treaty that brings together 192 countries to reduce greenhouse gases emissions [19].

Biofuels are thought to significantly contribute to a greener environment, and started to play a major role in the transportation sector. Bioethanol is considered the primary liquid fuel alternative because it can be blended with normal gasoline,

and is compatible with over 80 % of nowadays automobile engines [1]. Bioethanol is environmentally friendly, and sustainable with a short life cycle of raw material. Greenhouse gases emissions are reduced by 86 % compared to normal gasoline when cellulosic ethanol is used as liquid fuel [20].

The first generation bioethanol plants have been technologically established for many decades, and are exploited at commercial scales mainly in Brazil and USA, which are the top ethanol producers in the world. First generation plants rely on crops, such as sugar cane or corn, that are also used in the food industry. The massive investments into US ethanol plants increased the corn demand dramatically causing its price to triple [21]. The food versus fuel debate limits the further development of first generation bioethanol plants.

In contrast, second generation technology utilizes agricultural wastes such as wheat straw, bagasse or corn stover. This feedstock has a lower purchasing price and eliminates the food versus fuel debate. After many successful laboratory and pilot scale experiments, companies started to invest into scaling up the technology. The largest energy company in Denmark, i.e. DONG Energy A/S, created Inbicon A/S, a biotechnology company that focuses on second generation bioethanol. During the United Nations Climate Change Conference from 2009 in Copenhagen, Denmark (COP15), Inbicon opened the largest demonstration scale second generation bioethanol plant in the world at that time, capable of processing $4\,\mathrm{t\,h^{-1}}$ of raw biomass [3]. The plant is situated in Kalundborg, Denmark, and is integrated with Asnæs power plant following the Integrated Biomass Utilization System (IBUS) principle for costs reduction [4]. The bioethanol plant receives steam and process water from the Asnæs plant, and returns lignin bio-pellets that are co-burnt with coal for steam production. The second generation bioethanol technology reached commercial reality in 2012 [3], and in October 2013 the first commercial scale plant was commissioned in Crescentino, Italy by Beta Renewables, another biotechnology company [22].

Figure 1.1 shows a worldwide distribution of second generation bioethanol plants. The main development areas are in the USA, Brazil, and Europe. There are three main biorefineries already producing cellulosic ethanol at commercial scale in USA: Abengoa Bioenergy, DuPont, and POET-DSM running mainly on corn stover as feedstock, which is natural since USA is the world leader in corn production. Brazil follows with two other commercial scale refineries: GranBio and Iogen Energy utilizing sugarcane bagasse. Brazil is a leader in sugarcane production and first generation biofuels. European companies focus on research, comprising many centers for biotechnology development and following a licensing business model. Beta Renewables and Inbicon are two major European competitors with plans to

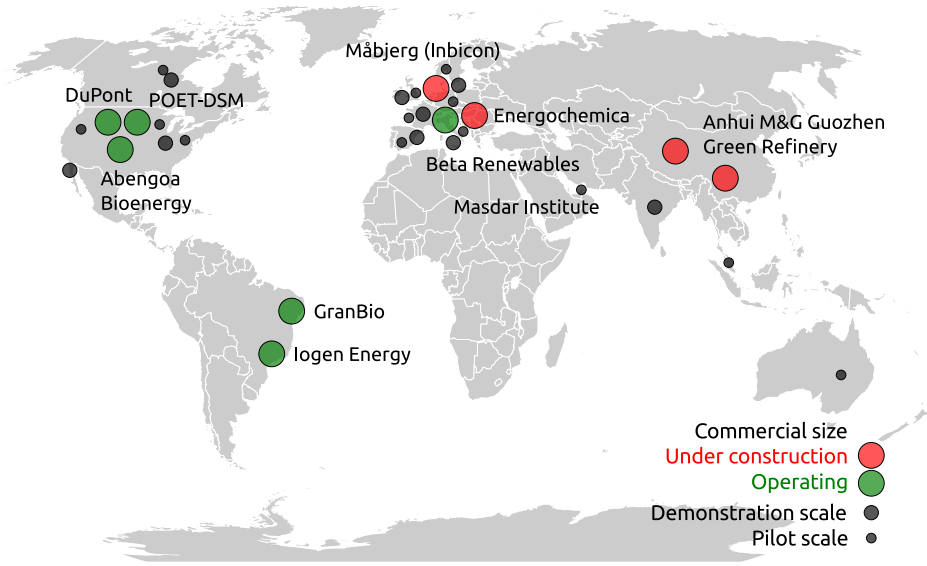


Figure 1.1: Second generation biorefineries around the world.

build commercial scale plants in Europe such as Måbjerg Energy Concept in Denmark (Inbicon) and Energochemica in Slovakia (Beta Renewables). Masdar Institute is also an important research center in cellulosic bioethanol situated in the Middle East where waste from palm trees is converted into biofuels. China plans to build commercial scale plants by licensing European technologies. Table 1.1 summarizes all commercial scale plants in operation nowadays and expected to open in the following years.

The biomass is transformed into biofuel through a series of conversion steps: pretreatment, enzymatic hydrolysis, sugar fermentation, and purification [4]. Figure 1.2 illustrates the focus of the Inbicon technology, i.e. a unique hydrothermal pretreatment with steam, the enzymatic hydrolysis with commercial Novozymes enzymes such as Cellic CTec, and C5 and C6 co-fermentation with Genetically Modified Organisms (GMO). The mechanical treatment and purification steps are borrowed from first generation plants with little adjustments. The major difference between generations is the feedstock type, i.e. agricultural wastes instead of crops.

Research work is also performed towards a third generation biofuel technology based on aquatic autotrophic organisms as feedstock, e.g. micro-algae or seaweeds [23]. Algae main characteristics are: rich in lipid content, easy to cultivate, and rapid growth rates, making them perfect candidates for biofuel production [24]. The third generation ethanol is at pilot scale stage at this time, and expected to flourish

in the upcoming years. The most advanced algae refinery is BioProcess Algae LLC located in USA, which is utilizing the CO_2 waste from a standard ethanol plant.

Biofuels are competitive in the liquid fuel market as long as oil prices are very high, or in the presence of government incentives and mandates [25]. Oil price reached minimum levels in 2009 and 2015. In such cases it is recommended to upgrade biorefineries to produce other chemicals with higher added value in order to remain competitive in a cheap oil environment [26]. For example, there is also a trend in research for adding value to lignin. In a second generation process, the remaining lignin from the feedstock could supply more than the entire energy for biofuel production [27]. Potential high-value products from isolated lignin include low-cost carbon fiber, engineering plastics and thermoplastic elastomers, polymeric foams and membranes, as well as a variety of fuels and chemicals all currently sourced from petroleum [28].

1.2 Motivation and Project Goals

The current research project has been financed by DONG Energy A/S in collaboration with the Danish Innovation Fund within the Industrial PhD program [29]. The main goal of the study is to improve and differentiate the Inbicon technology from its competitors by developing supervisory scientific tools for advanced control, optimization, and monitoring following a model-based design methodology.

Most industrial companies rely on static statistical models for improving the process. Such mathematical models are simpler and easier to develop but have many other disadvantages: offer no insights into process understanding as the model is entirely empirical, contain no time dynamics necessary for control design and

Table 1.1: Commercial scale second generation cellulosic ethanol plants in operation nowadays and expected to open soon [2].

Company	Location	Year
Beta Renewables	Italy	2013
Abengoa Bioenergy	USA	2014
POET-DSM	USA	2014
GranBio	Brazil	2014
DuPont	USA	2015
Iogen Energy	Brazil	2015
Måbjerg Energy Concept (Inbicon)	Denmark	2017
Energochemica	Slovakia	2017
Anhui M&G Guozhen Green Refinery	China	-

monitoring purposes, a small change into the experimental setup can invalidate the model, and sensitivity and uncertainty analysis to assess prediction reliability is difficult.

Another aspect is the process scale-up challenge. Very often empirical models developed at laboratory scale do not resemble the process behavior at a higher scale, i.e. demonstration or commercial. When scaling up a process, uniformity assumptions can fail, i.e. transport phenomena appear [30] creating temperature and pH gradients in large reactors. Another problem at higher scales is the delay time between tanks, also related to possible plug flow phenomena, which cannot be neglected anymore in control design. All published data by the time of this project dealt only with laboratory scale measurements, and all available models were developed based on small scale experiments.

In contrast, this work employs first principle dynamic modeling for industrial scale plants. It takes much more work to build such models than purely statistical methods and requires a more solid understanding of the root causes of the phenomenon but it is potentially more accurate [31]. The model library is then exploited to run dynamic simulations, design an optimization layer, formulate advanced control strategies, and construct real-time monitoring tools. The relation between the model and the scientific tools from this work is shown in Figure 1.3. Diagnosis and plant design methods are sown as possible extensions of the current work.

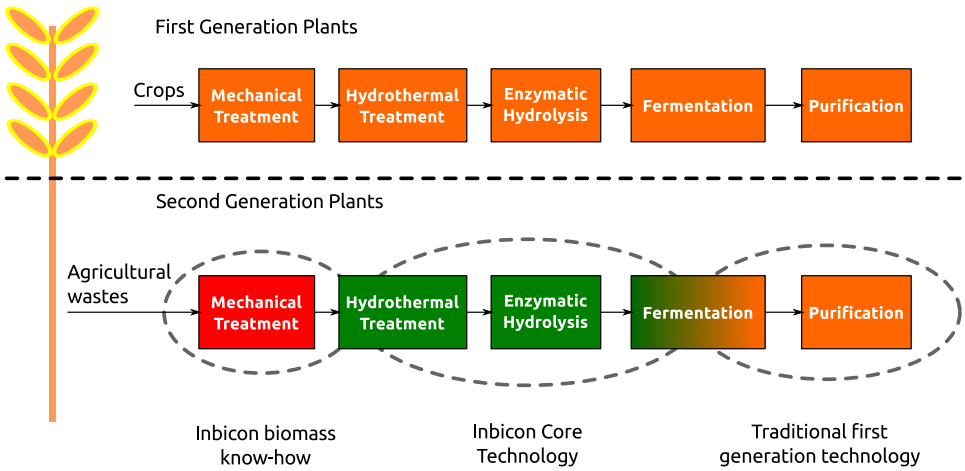


Figure 1.2: First and second generation biorefineries. The Inbicon technology is focused on hydrothermal pretreatment, enzymatic hydrolysis, and partially on C5 and C6 co-fermentation with GMO. Mechanical treatment and purification are borrowed from first generation plants.

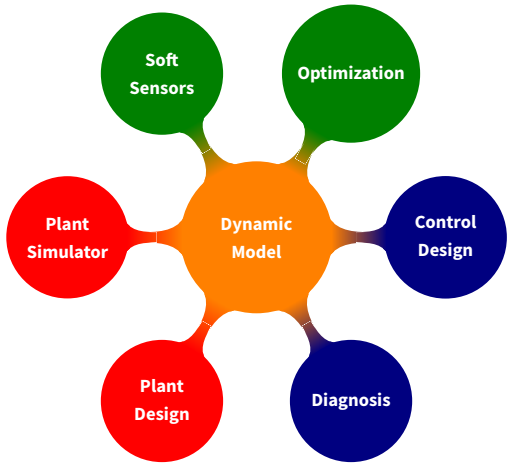


Figure 1.3: Dynamic modeling at the core of model-based scientific tools.

The specific goals of this research are:

1. To build a dynamic modeling library for pretreatment, enzymatic hydrolysis and fermentation. The library is designed to be modular in order to allow the user to change the configuration or add various components to the plant architecture. The models are validated against real data extracted from the Inbicon demonstration scale plant, and their reliability is assessed through a comprehensive sensitivity and uncertainty analysis.

2. To analyze the biorefinery in an integrated manner for establishing an overall optimal operation. The bioethanol production process consists of several sub-processes. Most studies analyze and optimize each conversion step individually in a decoupled manner without taking into account the trade-offs between stages. Most often the found optimal operation is in fact sub-optimal when comparing to the optimal point of an integrated system.

3. To explore the dynamic nature of the model to allow improved control design: the low level closed loop controls could be better tuned, or more advanced controllers such as model predictive and optimal control could be added to the overall plant automation layout. The objective is that this would be achieved in a simulation environment before real implementation.

4. To exploit the model to construct soft-sensors and state estimators for monitoring variables of interest that are difficult to obtain in real-time in reality. Biomass composition is measured in laboratory based on sample extractions either with Near Infrared (NIR) or High-Performance Liquid Chromatography

(HPLC) instruments. Such operations are time consuming posing delays of hours till the results are accessible. In contrast, a state observer can estimate the biomass composition in real-time at any point in the process, and improve prediction when measurements become available.

1.3 Thesis Outline

The thesis is written as a collection of publications. The main body text contains introductory principles and main research results, while the entire methodology is detailed in the appended articles. The thesis starts with a summary of all main contributions presented in Chapter 2. The dynamic modeling library for large scale biorefineries follows in Chapter 3. Only pretreatment and enzymatic hydrolysis are of interest since they constitute the core of the Inbicon technology. However, the modeling library is completed with a co-fermentation and distillation model developed in collaboration with DTU as a separate project, which is not included in this thesis. The approach and analysis from the modeling chapter creates a complete framework suitable for large scale processes that can be reiterated for any other complex system.

The optimization study comes next in Chapter 4. Trade-offs between conversion steps are identified, as well as the advantages of an optimization layer over a traditional operation based on a fixed recipe in liquefaction and fermentation. The optimizer provides setpoints for pretreatment temperature, enzyme dosage and yeast seed. The sensitivity analysis of the pretreatment model shows that reaction temperature has a large impact on downstream processes. At the same time, enzymes and GMO yeast have the highest costs in biorefinering, and it would be beneficial to save on these.

Enzymatic activity is highly sensitive to pH as indicated by the enzymes manufacturers. Advanced methods for controlling the key process parameters, i.e. pH in liquefaction and pretreatment temperature, are presented in Chapter 5. The thesis ends with conclusions and a list of perspectives and future ideas that could follow this work.

Chapter 2

Summary of Main Contributions

Journal Articles

The contributions of this research related to the modeling library and the optimization layer have been published (at the time of thesis submission, the optimization paper was undergoing the peer-review process) in three journal articles that were included in this thesis as appendices A, B and C:

(A) R. M. Prunescu, M. Blanke, J. G. Jakobsen, and G. Sin. "Dynamic modeling and validation of a biomass hydrothermal pretreatment process - A demonstration scale study". *AIChE Journal* (2015). DOI: 10.1002/aic.14954.

This study publishes for the first time a dynamic model for hydrothermal pretreatment with steam that is validated against demonstration scale real measurements. The model embeds mass and energy balances together with computational fluid dynamics for describing a large scale thermal reactor for biomass pretreatment. A comprehensive model analysis follows for assessing its sensitivity and uncertainty with respect to both feed and kinetic parameters. The dynamic trends of the process are well captured making the model suitable for developing advanced control and monitoring strategies for large scale plants. As an application of the model, the study includes the development of a state observer for estimating biomass components that are difficult to measure in reality.

(B) R. M. Prunescu and G. Sin. "Dynamic modeling and validation of a lignocellulosic enzymatic hydrolysis process - A demonstration scale study". *Bioresource Technology* 150 (Dec. 2013), pp. 393–403. DOI: 10.1016/j.biortech.2013.10.029.

This work formulates a complex dynamic model for enzymatic hydrolysis of cellulosic and hemicellulosic fibers suitable for large scale liquefaction reactors. The model includes a competitive conversion scheme for sugar production that was extended from previous works with hemicellulose hydrolysis, pH and viscosity calculators, and pH dependency on reaction kinetics. For the first time, model predictions are compared against demonstration scale real data extracted from the Inbicon plant. A sensitivity and uncertainty analysis is also performed to study modeling bottlenecks and for identifying the sensitive variables that affect most the uncertainty of model predictions.

(C) R. M. Prunescu, M. Blanke, J. G. Jakobsen, and G. Sin. "Model-Based Plantwide Optimization of a Large Scale Lignocellulosic Bioethanol Plant". *Submitted to AIChE Journal* (2015).

The scientific novelty of this work is the design of a plantwide model-based optimization layer for a large scale biorefinery. The objective is to maximize the economic profit by searching for the best trade-off between the conversion steps. The optimization solver can be triggered whenever there is a change in prices or feedstock composition, adapting the plant to market and operation conditions in order to maximize profitability at any given time. The optimization layer undergoes a sensitivity and uncertainty analysis for finding the variables that affect most the optimal point, and hence the economical profit. It is found that the optimization strategy is capable of reducing the uncertainty on the profit curve when compared to a traditional operation, and also allows running the plant in a wider nominal range with small impact on profitability. Feedstock composition impacts more on profit than model kinetics showing the need of accurate measurements of its composition.

Peer Reviewed Conference Proceedings (Web of Science)

The results concerning the advanced adaptive control strategies for process key parameters were disseminated in two peer-reviewed IEEE conference papers that were included in this thesis as appendices D and E:

(D) R. M. Prunescu, M. Blanke, and G. Sin. "Modelling and L1 Adaptive Control of Temperature in Biomass Pretreatment". *Proceedings of the 52nd IEEE Conference on Decision and Control*. Florence, Italy, 2013, pp. 3152–3159.

It has been shown in the sensitivity analysis of the pretreatment process that thermal conditions impact all downstream processes. Maintaining a steady reaction temperature, as well as quick reference tracking as imposed by the

optimization layer is of interest in this study. The main contribution refers to the application of an L1 adaptive output feedback controller for this type of process. The tuning method is also new consisting of numerical optimization for minimizing the Integral Absolute Error (IAE) cost function with respect to the controller parameters.

(E) R. M. Prunescu, M. Blanke, and G. Sin. "Modelling and L1 Adaptive Control of pH in Bioethanol Enzymatic Process". *Proceedings of the 2013 American Control Conference*. Washington D.C., USA, 2013, pp. 1888–1895.

Enzymatic activity is sensitive to the pH of the medium. The titration curve is highly nonlinear and poses a difficult challenge for any control strategy. The contribution from this work refers to the application of an L1 adaptive output feedback controller for enzymatic pH. The tuning method is new for this kind of processes, and relies on closed loop transfer function analysis that takes into account the interactions between the output predictor, the control signal filter and adaptation law.

Unpublished Work

There are two unpublished contributions included in this thesis:

(Section 3.4.4) A fast pH calculator with guaranteed accuracy for dynamic simulations:

pH is a key process parameter both in enzymatic hydrolysis and fermentation. The novelty from this section refers to a pH calculator that converges in a known amount of steps depending on the demanded accuracy. The algorithm is based on the charge balance of the liquid phase, and uses a modified bisection method that advances in logarithmic space for finding the pH level. The dynamic nature of simulations is also exploited, taking into account the solution from the previous simulation step in order to find tight bounds around the possible solution. The pH calculator has proven to be reliable and fast with guaranteed accuracy.

(Section 5.4) An optimal controller for the feed profile in glucose fermentation:

Fermentation reactors have a large volume and it can take days to fill the tanks till the desired hold-up, time when reactions

already take place. The contribution from this section shows how to compute an optimal feed rate profile such that inhibitors accumulation is avoided and yeast seed is minimized. The profile is found by formulating an Optimal Control Problem (OCP) and then compared to a classical constant feed strategy. The greatest benefit of a variable feed rate is that yeast amount is significantly reduced contributing to lower costs in fermentation.

Conference Presentations

All contributions were also disseminated in prestigious conferences through specialized session talks:

- R. M. Prunescu, M. Blanke, J. G. Jakobsen, and G. Sin. "Dynamic Modeling, Advanced Control, Diagnosis and Optimization of Large-Scale Lignocellulosic Biorefineries". *Proceedings of the AIChE 2015 Annual Meeting*. Salt Lake City, UT, USA, 2015.

- R. M. Prunescu, M. Blanke, J. G. Jakobsen, and G. Sin. "Plantwide Model-Based Optimization of a Large Scale Second Generation Biorefinery". *Proceedings of the AIChE 2015 Annual Meeting*. Salt Lake City, UT, USA, 2015.

- R. M. Prunescu, M. Blanke, J. G. Jakobsen, and G. Sin. "Model-Based Filtering of Large-Scale Datasets - A Biorefinery Application". *Proceedings of the AIChE 2014 Annual Meeting*. Atlanta, GA, USA, 2014.

- R. M. Prunescu and G. Sin. "Dynamic Simulation, Sensitivity and Uncertainty Analysis of a Demonstration Scale Lignocellulosic Enzymatic Hydrolysis Process". *Proceedings of the AIChE 2014 Annual Meeting*. Atlanta, GA, USA, 2014.

- R. M. Prunescu, M. Blanke, and G. Sin. "Advances in Monitoring, Diagnosis and Control of Biorefineries". *Proceedings of the 9th World Congress of Chemical Engineering*. Seoul, South Korea, 2013.

Chapter 3

Dynamic Modeling and Analysis

3.1 Introduction

This chapter presents the main results from two scientific journal publications included in appendix as Paper A and Paper B. The focus is placed on biomass pretreatment and enzymatic hydrolysis, which are the core processes of the Inbicon technology. The co-fermentation model constitutes the subject of a separate project, while the purification technology is state of the art with no customization for Inbicon. Therefore the liquefaction and distillation processes are not included in this work.

The chapter starts with a detailed description of the Inbicon second generation bioethanol plant, followed by the model analysis methodology. The entire mathematical model library is then summarized, and the main sensitivity and uncertainty analysis results are discussed. The chapter ends with conclusions and suggestions for future modeling improvements and maintenance.

3.2 Process Description

This section describes the composition of feedstock used in second generation biorefineries including technical details for each biomass conversion step, i.e. pretreatment, enzymatic hydrolysis, co-fermentation and purification.

3.2.1 Biomass Characterization

Lignocellulosic biomass consists of cellulose, hemicellulose, lignin, ash, and other residues in negligible amounts [37]. Hemicellulose further divides into xylan, arabinan, galactan, mannan and acetyl groups [38]. Table 3.1 shows different biomass compositions depending on agricultural waste type, e.g. bagasse, wheat straw, miscanthus, corn stover, or quinoa stalks. Even if the biomass is of the same

type, it can still have a different percentage distribution of its components due to seasonality, harvest location, amount of rain, or used fertilizers for growing the crops. The solid composition of biomass is typically measured with NIR instruments based on laboratory samples, and is essential for determining the biofuel potential for each biomass type [39]. The feedstock has an initial dry matter of over 85 %.

Table 3.1: Composition of different raw biomass types in percentage of dry matter. Some components were not measured (n/m).

Biomass	Cellulose	Xylan	Arabinan	Lignin	Ash	Acetyls	Other	Reference
Wheat straw	39.8	24.5	2.8	22.6	4.2	n/m	6.1	[40]
Wheat straw	42.1	23.4	1.9	21.6	2.1	n/m	8.9	[41]
Corn stover	35.0	18.5	n/m	13.9	3.8	3.2	25.6	[42]
Corn stover	34.0	19.2	2.5	12.3	4.7	2.9	24.4	[43]
Bagasse	39.0	21.8	1.8	24.8	3.9	3.3	5.4	[43]
Miscanthus	38.2	20.9	1.5	26.4	2.6	4.1	6.3	[44]
Quinoa stalks	35.7	15.4	3.5	21.9	4.2	2.7	16.6	[38]

3.2.2 The Inbicon Biorefinery

The Inbicon biorefinery is integrated with Asnæs power plant (also owned by DONG Energy A/S) following the IBUS principle [4]. The symbiosis between the refinery and the plant is illustrated in Figure 3.1. The biorefinery receives steam at 18 bar for a low cost, and returns lignin-pellets to be co-burnt with coal in the power plant for steam and power generation. The steam fuels the biomass pretreatment and purification processes. Figure 3.1 illustrates the first version of Inbicon where the C5 sugars from the pretreatment process were transformed into molasses along with wasted yeast from fermentation, and sold to local farmers to feed their cattle. The conversion steps are: pretreatment, enzymatic hydrolysis or liquefaction, fermentation, and purification. Bioethanol is the main refined product, followed by two other by-products, i.e. C5 molasses and lignin-pellets. The first version of the Inbicon demonstration scale plant produces 576 kg of biofuel, 1484 kg of molasses, and 1740 kg of lignin-pellets from 4 t of dry biomass, which is the nominal throughput per hour of the plant [3]. The amount of lignin-pellets generates enough energy to cover the requirements for biofuel production [27].

The second version of Inbicon is shown in Figure 3.2. The C5 sugars are no longer transformed into molasses but rather used in co-fermentation with GMO yeast. The stream with C5 sugars by-passes the enzymatic hydrolysis tanks and is directed to fermentation. Results show that the latest Inbicon version increases the

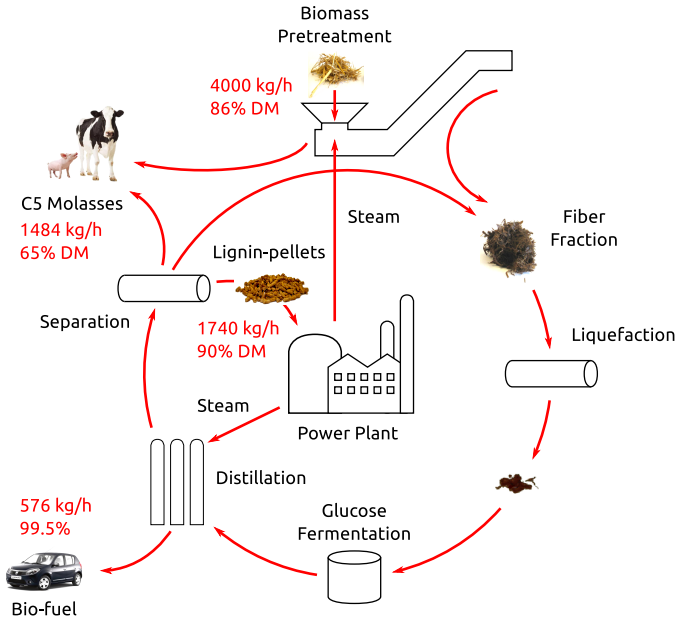


Figure 3.1: The integration of the first version of Inbicon with a local power plant following the IBUS principle. The characteristics of the first version are: glucose fermentation, while the C5 sugars are sold as molasses to local farmers.

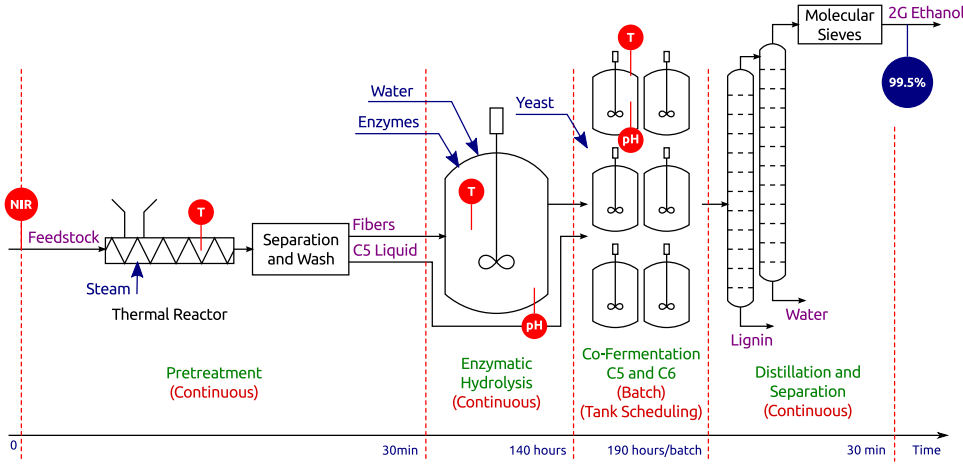


Figure 3.2: The second version of Inbicon. Instead of selling C5 sugars as molasses, they are co-fermented with C6 sugars using enhanced GMO yeast for ethanol production.

ethanol yield by 40 % to approximately 806 kg per 4 t of dry straw [45]. The same conversion steps apply as in the first version.

3.2.3 Pretreatment

Biomass pretreatment is the key process to unlocking low-cost cellulosic ethanol [46]. Lignin or the wooden part supports the plant structure, and also protects the cellulosic fibers. The scope of the pretreatment process is to open the biomatrix by relocating the lignin, and to partially hydrolyze the hemicellulose in order to expose cellulosic fibers for the enzymatic process downstream [40]. There are various methods of lignocellulosic biomass pretreatment: hydrothermal, autohydrolysis, acid or alkaline hydrolysis [44], SO_2-catalyzed steam pretreatment [47] and organosolv processes [48]. The most cost effective and environmentally friendly method is the hydrothermal pretreatment process, which requires no additional chemicals [49], and has already been tested at larger scales by many companies, such as Inbicon.

The Inbicon pretreatment process is continuous, and occurs in a long cylindrical thermal reactor pressurized with saturated steam from the nearby power plant. The biomass is pushed horizontally from inlet to outlet with a constant speed through a steam atmosphere. Biomass pretreatment is sensitive to retention time and temperature [50], which can be adjusted accordingly. Typical temperatures lie between 160 °C to 210 °C with retention times between 5 min to 15 min. A longer retention time requires a lower temperature, and vice-versa. A constant horizontal speed translates into a constant throughput or retention time, which is desired in a large scale facility. The reactor is equipped with a controller for setting the reaction temperature by changing the pressure inside the thermal reactor [12]. The pressure is controlled by injecting saturated steam through the bottom of the tank near the inlet from a fresh steam supply pipe. Before entering the reactor, the biomass is brought to the corresponding pressure by a particle pump, which pressurizes an amount of biomass with recycled steam from the thermal reactor every 2 min.

Hemicellulose hydrolysis creates the following by-products during pretreatment: xylose and xylooligomers, which are important to predict because they represent the C5 sugars yield in co-fermentation, and are also strong liquefaction inhibitors [51]; sugar degradation products such as furfural and 5-HMF, which inhibit the fermentation process [52]; organic acids, e.g. acetic, succinic or lactic acid, which disturb the pH of the medium in liquefaction and fermentation [6]; and pseudo-lignin, a spherical droplet with lignin like structure, created by the reaction between inhibitors and carbohydrates [53]. These facts show that the pretreatment products affect all downstream processes.

After pretreatment, the biomass is washed and separated into solid and liquid

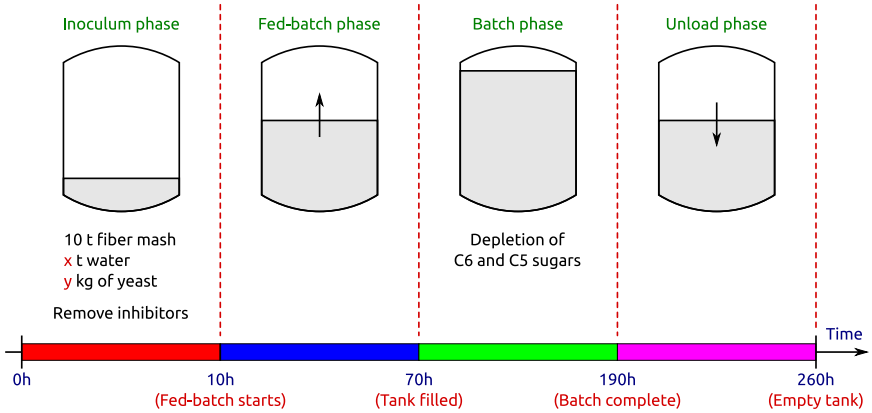
parts by a screw press. The solid part is rich in cellulose, while the liquid part contains the C5 sugars that were produced due to hemicellulose hydrolysis.

3.2.4 Enzymatic Hydrolysis

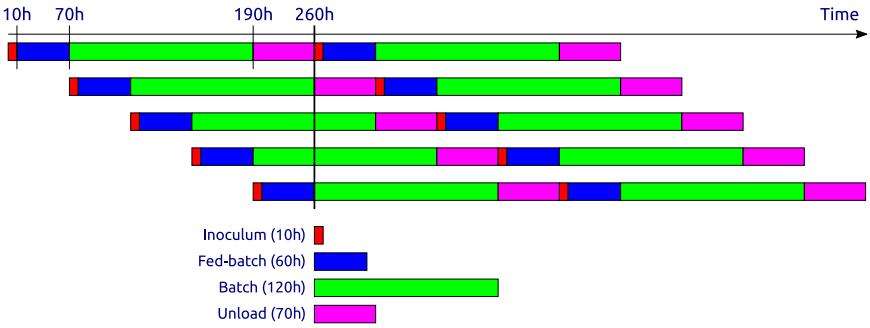
A conveyor belt transports the cellulosic fibers to the enzymatic hydrolysis reactor. The Inbicon liquefaction process is also continuous, and occurs in several reactors connected in series. The first tank is a 5-chambers hydrolysis reactor presented in [54] that was specifically designed for high dry matter liquefaction of biomass preferably around 35 % [55]. There is an abrupt change in viscosity in the first hours of hydrolysis allowing the slurry to be easily pumped afterwards. The rheology phenomena is well documented in torque measurements, which decrease exponentially [56]. The following tanks are conventional Continuous Stirred Tank Reactors (CSTRs) linked in series to meet the necessary hydrolysis time of 140 h. There are many commercially available enzymes, e.g. Cellic CTec2 [57], Cellic CTec3 [58], Cellic HTec3 [59]. The hydrolysis retention time can be adjusted either by changing the tank hold-ups (preferably) or by setting a different refinery throughput.

Nowadays enzymes are capable of hydrolyzing both cellulosic and hemicellulosic fibers. Cellulose hydrolysis produces glucose with cellobiose intermediate product, while hemicellulose hydrolysis is more complex leading to xylooligomers, xylose and organic acids production. The enzymatic mixture is a cocktail of cellulase and xylanase. Cellulase hydrolyzes cellulose, and consists of exo-β-1.4-cellobiohydrolase, endo-β-1.4-glucanase, and β-glucosidase [60]. Cellulose is a long solid polymer or chain made up of glucose units. The endo-β-1.4-glucanase randomly breaks internal bonds from the cellulosic fibers creating new chain ends. The exo-β-1.4-cellobiohydrolases further cleave the endo-glucanase products producing cellobiose. The β-glucosidase enzymes breaks cellobiose into glucose. The xylanase enzymes behave similarly with xylooligomers intermediate product but along with hemicellulose hydrolysis it also releases acetyls, which produces acetic acid affecting the pH of the medium in the reactor.

The enzymatic activity is sensitive to pH and temperature following a bell shaped efficiency curve with a single peak [11]. Temperature control is easily achieved, while pH control has many challenges due to the nonlinearities in the titration curve [11]. All enzymatic hydrolysis reactors have temperature and pH controllers. The degree of biomatrix opening or treatment severity, and some pretreatment inhibitors, i.e. xylooligomers and xylose, reduces the enzymatic activity. Also, liquefaction is a competitive mechanism with product inhibition. Enzymes can be inhibited but also irreversibly deactivated in time and due to wrong temperature exposure [61].



(a) Fermentation process consisting of 4 phases: inoculum (10 h), fed-batch (60 h), batch (120 h), and unload (70 h).



(b) An example of a fermentation process with 5 scheduled reactors. The reactors are scheduled such that the liquefied fiber inflow and ethanol outflow stay constant with minimum interrupts.

Figure 3.3: Fermentation process: phases and scheduling.

3.2.5 Fermentation

The C5 liquid part from pretreatment and the C6 sugars from the enzymatic hydrolysis are mixed in large tanks that have a maximum capacity of 250 m^3. The C5 and C6 co-fermentation process has four stages illustrated in Figure 3.3(a): inoculum, fed-batch, batch, and unload. In the inoculum phase 10 t of hydrolyzed fibers are mixed with yeast and water for an initial biomass growth. The inoculum phase ends when all inhibitors are consumed in maximum 10 h. The fed-batch stage then starts for filling the entire fermentation reactor with hydrolyzed fibers and C5 liquid fraction from pretreatment. It can take up to 60 h to reach the desired reactor hold-up, time when reactions already take place. Once the tank is filled, the batch step commences. Both C5 and C6 sugars are depleted at this stage in about 120 h. The reactor is then unloaded and the stream is directed to purification.

The fermenters have temperature and pH controllers. Typical operation conditions are 35 °C and 5.5 pH units, which are optimal for the GMO yeast. The enzymes are still active during fermentation making the system a simultaneous saccharification and fermentation process. The liquefied fibers still contain solid cellulose and hemicellulose that were not entirely hydrolyzed during liquefaction because of product inhibition. As sugars are depleted in fermentation for ethanol production, the sugars inhibitory effect decreases, and enzymes continue the liquefaction process simultaneously.

In large scale facilities, fermentation runs in a batch manner requiring more reactors to run in parallel according to a scheduling algorithm as illustrated in Figure 3.3(b). The operation is aligned such that input and output streams flow continuously with minimum interruptions.

3.2.6 Purification

The first distillation column separates lignin from the stream. The lignin is sent to a local evaporation unit, which creates the lignin bio-pellets. The second distillation column purifies ethanol further, which reaches 99.5 % purity after the molecular sieves. The bioethanol is stored in underground tanks till an oil company transports them to their facilities to be blended with regular gasoline. The bioethanol is sold at petrol stations as E10, E15, E20, or E85, the number indicating the percentage of ethanol from the mix, e.g. E10 contains 10 % bioethanol and 90 % gasoline.

3.3 Model Analysis Framework

The model analysis framework is summarized in Table 3.2, which is detailed in the next 7 steps:

1. Formulate the mathematical model structure as a system of nonlinear Ordinary Differential Equations (ODEs). Identify states, inputs, outputs and model parameters:

$$\begin{aligned} \dot{x} &= f(x,u,\theta) \\ y &= g(x,\theta) \end{aligned} \tag{3.1}$$

   f is an array of nonlinear functions of states x, inputs u, and parameters θ. The outputs y are defined as nonlinear functions g of states x and model parameters θ.

2. The second step is to calibrate the model considering the entire set of parameters. This system identification exercise follows the nonlinear least squares method for grey-box models, which should give the set of parameters that has

the smallest sum of squared errors between model predicted output and actual measurements [62]:

$$\min_{\theta} \sum_{i=1}^{N} e_i^2 \tag{3.2}$$

e_i is the estimation error at sample time i defined as $e_i = y_i - \hat{y}_i$, the real measurement y_i and the predicted output $\hat{y}_i$. In the present case, this is a nonlinear least squares problem and local minima can be obstacles.

3. The third step is to investigate which model parameters could be identified given the input and the model structure [63]. This selection is achieved through assessment of sensitivity of the partial derivatives of the cost function with respect to each model parameter. The delta mean square δ_{ik}^{msqr} defined by [64] is a measure of sensitivity suitable for time varying signals:

$$\delta_{ik}^{msqr} = \sqrt{\frac{1}{N} s_{nd,ik}^{\top} s_{nd,ik}} \tag{3.3}$$

where k is the parameter index, i is the model output index, N is the number of samples, and $s_{nd,ik}$ is a vector with the non dimensional sensitivity calculated in each sample:

$$s_{nd,ik} = \frac{\partial y_i}{\partial \theta_k} \frac{\theta_k}{sc_i} \tag{3.4}$$

$\partial y_i / \partial \theta_k$ represents the output variation with respect to parameter θ_k, and sc_i is a scaling factor with the same physical dimension as the corresponding observation in order to make this measure non dimensional. In this study, the scaling factor is chosen as the mean value of output i:

$$sc_i = \frac{1}{N} \sum_{1}^{N} y_i(k) \tag{3.5}$$

After computing the sensitivities δ^{msqr}, all parameters are ranked with respect to their value of δ^{msqr}. Parameters that have low sensitivity are more uncertain than those with high sensitivity and would not contribute to model accuracy. In case of systems with multiple outputs, a cumulative sensitivity measure is defined as:

$$\delta_{k}^{msqr} = \sum_{i=1}^{n_y} \delta_{ik}^{msqr} \tag{3.6}$$

The relevant subset of parameters is selected based on δ^{msqr} being higher than a threshold.

4. The reduced set of parameters is properly estimated following the same minimization technique from step 2. In this case, the real measurements are

in fact NIR or HPLC laboratory datasets from the demonstration plant. The whole dataset is split into estimation and validation subsets. The parameter estimation procedure runs on the estimation dataset. The correlation matrix and standard deviations of the estimates are also calculated.

5. This step quantifies the prediction uncertainty. Having the covariance matrix and standard deviations from the previous step allows LHS with correlation control [65]. The feed parameters is another source of uncertainty and is included in this analysis. Monte Carlo simulations are then run with sampled parameter values and the 5th-50th-95th percentiles of the model predictions are found.

6. A global sensitivity analysis follows by fitting a linear model from parameters to model predictions from the Monte Carlo simulations [66, 67]:

$$y_{reg_i} = a + \sum_k b_k \theta_k \tag{3.7}$$

where y_{reg_i} is the i^{th} output, and a and b_k are the linear model parameters. The standardized regression coefficients β are a global sensitivity measure, and are defined as:

$$\beta_k = \frac{\sigma_{\hat{\theta}_{R_k}}}{\sigma_{y_i}} b_k \tag{3.8}$$

where β_k is the β coefficient, $\sigma_{\hat{\theta}_{R_k}}$ is the standard deviation of the parameter estimate, σ_{y_i} is the standard deviation of output i, and b_k is the linear model parameter. β_k is an indicator for how much the parameter uncertainty contributes to the prediction uncertainty.

7. The model estimation error or the residuals are analyzed in this step. A simulation is run with the estimated parameters using the entire set of data (both validation and estimation sets). The residuals distribution and autocorrelation are calculated in order to assess the quality of model predictions. A good model captures most of the signal in measurements and is characterized by residuals being Gaussian with uncorrelated increments.

The sensitivity and uncertainty analysis based on Monte Carlo simulations has been successfully applied in numerous situations, e.g. in enzymatic biodiesel production [68], cellulose hydrolysis [8], wastewater plant treatment [66], or lignocellulosic ethanol plants [9].

Table 3.2: Model analysis methodology.

#	Step	Description	Output
1	Model formulation	Define model structure through ODEs; Define model outputs; Define model parameters; Define model states; Define model inputs.	$\dot{x} = f(x,u,\theta)$ $y = g(x,\theta)$ θ x u
2	Model initialization	Initialize all model parameters to obtain a good working model fit;	θ_0
3	Sensitivity analysis	Calculate sensitivity measure; Find an identifiable parameter subset.	δ^{msqr} θ_{R_0}
4	Parameter estimation	Identify parameter subset; Correlation matrix; Standard deviation for parameters.	$\hat{\theta}_R$ $\mathbf{R}_\theta$ σ
5	Uncertainty analysis	Calculate model prediction uncertainty through Monte Carlo simulations;	5th-95th percentile
6	Global sensitivity analysis	Standardized regression coefficients.	β
7	Residual analysis	Run simulation with the estimated parameters and using the entire dataset; Check probability distribution of model estimation errors or residuals; Compute the autocorrelation function.	

3.4 Mathematical Model Development

This section presents the dynamic model library. All models are based on mass and energy balances with complex biomass conversion routes. The models are grouped by processing step, i.e. pretreatment, liquefaction and co-fermentation. This study does not include the purification stage due to the fact that the distillation part is state of the art technology with no Inbicon customization.

3.4.1 Pretreatment

The pretreatment thermal reactor is illustrated in Figure 3.4 where the steam and biomass layers are distinguished. The steam layer is considered uniform and is treated as a single unit. The steam atmosphere has been modeled in Paper D where

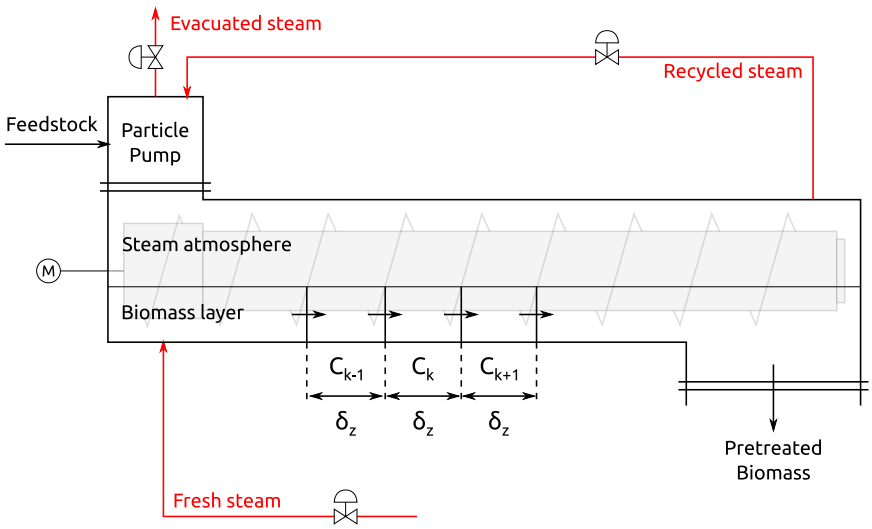


Figure 3.4: Biomass and steam create two distinct layers in the thermal reactor. The biomass layer employs computational fluid dynamics tools for modeling, while the steam layer is treated as a single cell.

the mass and energy balances are established [12]. The steam state variables are: steam mass and enthalpy (energy). The total mass of biomass is tracked through another state variable such that the empty volume from the reactor can be calculated. Steam expands in empty volume with very fast dynamics, therefore it is considered a cell with uniform properties. If the steam is saturated then one parameter is enough to derive any other steam property. If the steam is superheated or wet then two parameters are required to derive any other steam property. All steam properties are calculated based on linear interpolation of tabular data defined by The International Association for the Properties of Water and Steam - Industrial Formulation 1997 (IAPWS IF97) [69].

The biomass layer is split into multiple cells using simplified Computational Fluid Dynamics (CFD) tools for modeling the composition and temperature profiles. Figure 3.4 illustrates this concept. The entire modeling methodology is described in Paper A. The CFD tools refer to the Convection Diffusion equation for describing the transport phenomena [70]:

$$\frac{\partial C}{\partial t} = -\nabla(vC) + \nabla(D\nabla C) + R \tag{3.9}$$

C is the variable of interest, i.e. the concentration of a certain chemical compound expressed in $\mathrm{g\,kg^{-1}}$ of slurry, ∇ is the gradient operator, v is the transport speed vector, D is the diffusion coefficient, and R is a vector with all reaction rates.

Several assumptions are made for simplifying Equation (3.9). The horizontal pushing speed is set constant in order to ensure a uniform throughput. A tank of 12 m with a retention time set to 15 min pushes the biomass with a speed of $0.013\,\mathrm{m\,s^{-1}}$, which is too low to mix the biomass vertically [5]. If the reactor filling level is low then vertical uniformity is assumed as in Paper A where the throughput is set to $1\,\mathrm{t\,h^{-1}}$. However, if the reactor level is high then vertical temperature gradients appear and the biomass is pretreated differently in layers of temperature as in [71]. The biomass consists of solids, which do not diffuse. The solubles that are created during hydrolysis have a low diffusion coefficient due to the physical barrier of the snail pallets that push the biomass along the reactor [5]. Also biomass has a high dry matter content of approximately 40 % inside the reactor. These facts lead to a domination of convective effects and the diffusion part of Equation (3.9) can be dropped, i.e. $D \approx 0$. The tank cylinder is longer in one direction, which makes it necessary to discretize in space only along one axis. Considering all these assumptions, Equation (3.9) is simplified to:

$$\frac{\partial C}{\partial t} = -v\frac{\partial C}{\partial z} + R \tag{3.10}$$

where z is the axis oriented along the reactor from inlet to outlet. Equation 3.10 is then discretized in space using a Upwind Difference Scheme (UDS), which is a more stable technique for moving environments [72]:

$$\frac{\partial C_k}{\partial t} = \frac{v}{\delta_z}\left(C_{k-1} - C_k\right) + R_k \tag{3.11}$$

C_k, C_{k-1} and δz are illustrated in Figure 3.4 and represent the central cell k with its western neighbor $k-1$, and the width of the discretization step δz. The western boundary conditions are of Dirichlet type calculated based on a virtual cell derived from known input biomass composition while the eastern boundary conditions are considered of Neumann type or zero material loss through the reactor shell. The methodology for establishing the boundary conditions is similar to the one derived in [71, 72].

Reaction vector R_k is calculated considering the conversion paths from Figure 3.5, which were extended from the original work in [73] with acid, xylooligomers, pseudo-lignin, and 5-HMF production. The detailed derivation of the reaction rates is included in Paper A. Xylan hydrolysis produces xylose with xylooligomers intermediate product. Xylose can be further degraded to furfural. Biomass has a reduced content of arabinan, which hydrolyzes completely to arabinose. Along with hemicellulose hydrolysis, acetyls are released into the medium forming acetic acid. To a lower extent, cellulose can also be hydrolyzed leading to glucose production,

which further degrades to 5-HMF. Carbohydrates and inhibitors react in order to create pseudo-lignin.

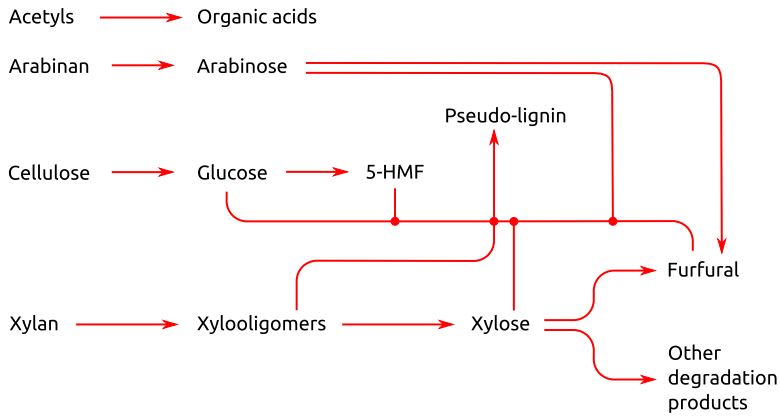


Figure 3.5: Hydrothermal pretreatment conversion paths [5].

The composition vector C_k and the reaction rate array R_k from Equation (3.11) are placed in vector form:

$$C_k = \begin{bmatrix} C_{C_S} & \text{(Cellulose)} \\ C_{X_S} & \text{(Xylan)} \\ C_{A_S} & \text{(Arabinan)} \\ C_{L_S} & \text{(Lignin)} \\ C_{Ac_S} & \text{(Acetyls)} \\ C_G & \text{(Glucose)} \\ C_{Xo} & \text{(Xylooligomers)} \\ C_X & \text{(Xylose)} \\ C_A & \text{(Arabinose)} \\ C_{Ac} & \text{(Acid)} \\ C_F & \text{(Furfural)} \\ C_H & \text{(5-HMF)} \\ C_W & \text{(Water)} \\ C_O & \text{(Other)} \end{bmatrix} \qquad R_k = \begin{bmatrix} -r_G \\ -r_{Xo} \\ -r_A \\ r_L \\ -r_{Ac} \\ r_G - r_{O_G} - (1-\alpha) r_{L_G} \\ r_{Xo} - r_X - (1-\alpha) r_{L_{Xo}} \\ r_X - r_{F_X} - r_{O_X} - (1-\alpha) r_{L_X} \\ r_A - r_{O_A} - r_{F_A} - (1-\alpha) r_{L_A} \\ r_{Ac} \\ r_F - \alpha r_{L_F} \\ r_H - \alpha r_{L_H} \\ 0 \\ r_{O_X} + r_{O_G} + r_{O_A} \end{bmatrix} \tag{3.12}$$

To ensure mass conservation the following relations have to stand true:

$$\sum R_k = 0\,\mathrm{g/(kgs)} \qquad \sum C_k = 1000\,\mathrm{g\,kg^{-1}} \tag{3.13}$$

All reaction rates are temperature dependent. Due to lack of agitation inside the reactor, temperature gradients appear along the reactor. The energy balance in the biomass layer is derived in a similar way based on the Convection Diffusion Equation.

Biomass has insulation properties resulting in a low heat diffusion coefficient. Only convective effects are assumed. The variable of interest in this case is the enthalpy h expressed in $\mathrm{kJ\,kg^{-1}}$:

$$\frac{\partial h}{\partial t} = -v\frac{\partial h}{\partial z} + Q_k \Rightarrow \frac{\partial h_k}{\partial t} = \frac{v}{\delta z}(h_{k-1} - h_k) + Q_k \tag{3.14}$$

where Q_k is the energy inflow in cell k. The same cell grid is used as in the biomass composition case. The steam injection occurs near the inlet and is lumped into the boundary conditions, which are detailed in Paper A. The temperature gradient is then obtained by dividing the enthalpy to the specific heat.

3.4.2 Enzymatic Hydrolysis

The model derivation of the enzymatic hydrolysis process is detailed in Paper B. A reactor with 5 chambers designed for high dry matter liquefaction is sketched in Figure 3.6. The tank has a common shaft with large pallets attached to it that ensure vertical uniformity. The transport phenomena employs CFD methods in a similar way as in the pretreatment case. Due to the initial high dry matter content of about 35 %, plug flow is assumed in the first hours of hydrolysis. The reactor is discretized along a single axis, i.e. the shaft or along the reactor. During nominal operational conditions the process is continuous with non zero inflow and outflow where convective effects dominate. However, it might happen due to upstream or downstream faults that the inflow and outflow stop. In such a case the system becomes a batch process with dominant diffusion effects for solubles such as glucose, xylose, acetic acid, and furfural. Therefore the diffusion coefficient is not dropped from the Convection Diffusion Equation, and appears as a function of viscosity.

Enzymes bound to solids in order to hydrolyze them but not entirely. There is an equilibrium between bounded and free enzymes that follows a Langmuir isotherm expression [60]:

$$\frac{C_{E_{i_B}}}{C_S} = E_{M_i}\frac{K_{A_i}C_{E_{i_F}}}{1 + K_{A_i}C_{E_{i_F}}} \tag{3.15}$$

where $C_{E_{i_B}}$ is the concentration of bounded enzymes of type i on solids C_S. E_{M_i} is the maximum adsorbed enzymes, and K_{A_i} is the Langmuir adsorption constant. $C_{E_{i_F}}$ is the concentration of free enzymes. Index i iterates through enzymes types, which are 4 in total: endo-exo cellulase, β-glucosidase, endo-exo xylanase, and β-xylosidase [35].

The reactions follow the conversion paths from Figure 3.7. Cellulase enzymes most often break the long solid chain of cellulose into smaller units like cellobiose with rate r_1. Cellobiose acts as an intermediate product for glucose production. Cellulase can also detach glucose units from the chain endings of polymers longer

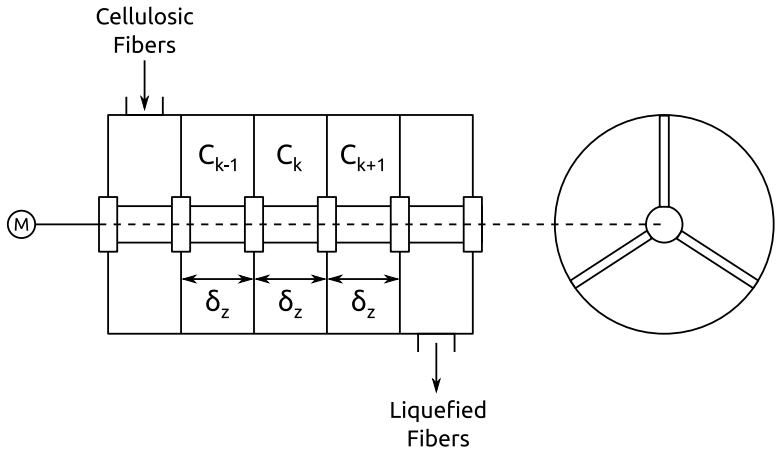


Figure 3.6: High dry matter liquefaction reactor [54].

than cellobiose leading to direct glucose production with rate r_2. β-glucosidase further decomposes cellobiose to glucose with rate r_3. Hemicellulose is hydrolyzed through a similar mechanism with xylooligomers intermediate product and acetic acid as by-product. Xylanase decomposes xylan into xylooligomers with rate r_4, and also straight to xylose when acting on the chain endings with rate r_5. Xylooligomers

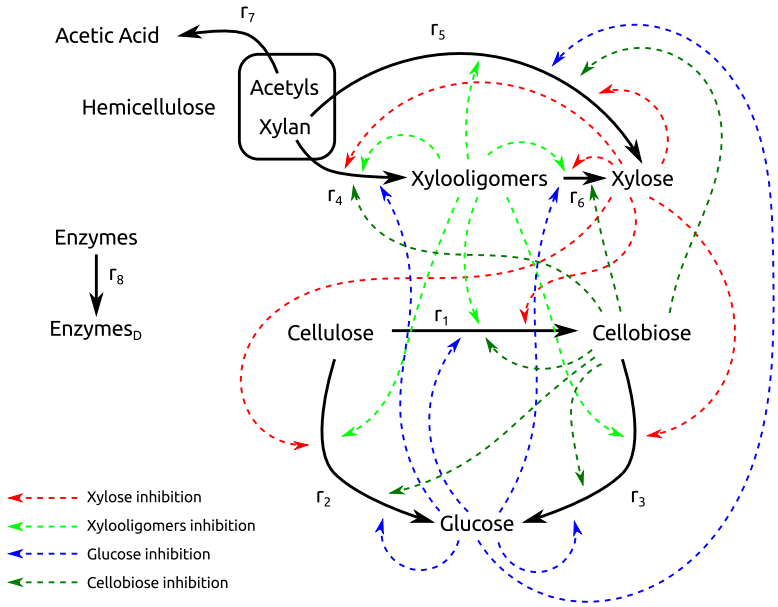


Figure 3.7: Enzymatic hydrolysis conversion mechanism with inhibition. Extended from [6] with xylooligomers intermediate product, acetyl groups, and enzyme deactivation.

are further hydrolyzed to xylose with rate r_6. Acetyls are part of the hemicellulose, and as the solid fibers are decomposed they get released in the medium forming acetic acid with rate r_7. Enzymes slowly deactivate irreversibly in time with rate r_8. The mathematical relations for all reaction rates are given in Paper B and in the supplementary material from Paper C where the following extensions are added: severity dependency of all reaction rates, enzymes deactivation, acid production from acetyl groups, and xylooligomers intermediate product in hemicellulose hydrolysis [7].

The conversion mechanism is competitive with product inhibition. The dashed lines from Figure 3.7 illustrate all possible inhibition phenomena due to xylose, xylooligomers, glucose and cellobiose. The inhibitory effects are not relevant for all reaction rates. The sensitivity analysis helps identify the important inhibition paths.

The composition vector C_k and the reaction rates array R_k are shown next:

$$C_k = \begin{bmatrix} C_{C_S} & \text{(Cellulose)} \\ C_{X_S} & \text{(Xylan)} \\ C_{L_S} & \text{(Lignin)} \\ C_{Ac_S} & \text{(Acetyls)} \\ C_{Ac} & \text{(Acid)} \\ C_C & \text{(Cellobiose)} \\ C_G & \text{(Glucose)} \\ C_{Xo} & \text{(Xylooligomers)} \\ C_X & \text{(Xylose)} \\ C_F & \text{(Furfural)} \\ C_H & \text{(5-HMF)} \\ C_B & \text{(Base)} \\ C_E & \text{(Enzymes)} \\ C_W & \text{(Water)} \\ C_O & \text{(Other)} \end{bmatrix} \quad R_k = \begin{bmatrix} -r_1 - r_2 \\ -r_4 - r_5 \\ 0 \\ -r_7 \\ r_1 - r_3 \\ r_2 + r_3 \\ r_4 - r_6 \\ r_5 + r_6 \\ r_7 \\ 0 \\ 0 \\ 0 \\ -r_8 \\ 0 \\ r_8 \end{bmatrix} \tag{3.16}$$

To ensure the mass balance Equation (3.13) also needs to stand true.

The model is extended with a pH and viscosity calculator. The pH is important for control purposes because enzymes are very sensitive to small deviations in pH. The viscosity indicates the state of the liquefaction process and is important for performance monitoring purposes. Viscosity is also easier to measure in reality in real-time than extracting samples and analyzing them in the laboratory. Viscosity decreases exponentially as solids are liquefied and increases linearly as sugars are dissolved in the liquid phase [6]. The pH charge balance and the viscosity calculator are elaborated in Paper B.

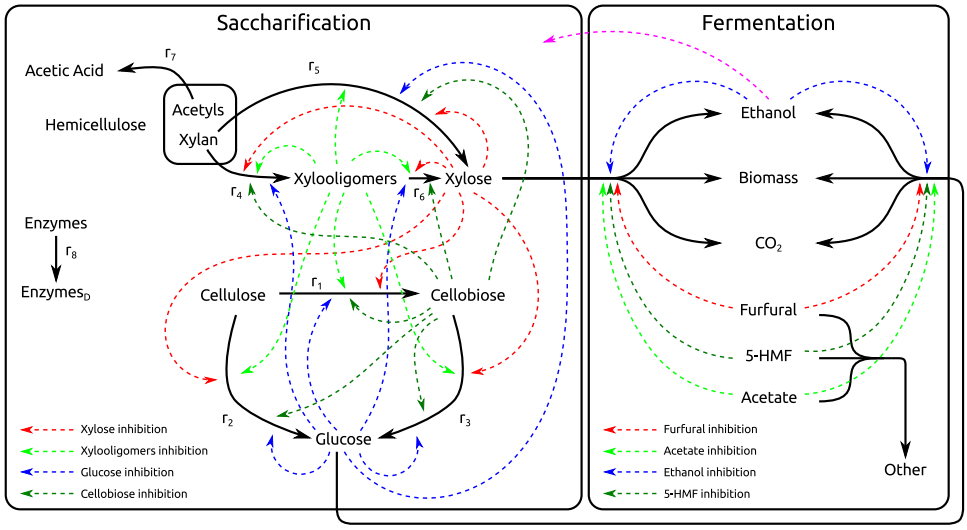


Figure 3.8: Simultaneous saccharification and fermentation.

3.4.3 Fermentation

The co-fermentation process is modeled as a Continuous Stirred Tank Reactor (CSTR) with the conversion paths from Figure 3.8. The saccharification part is identical to the enzymatic hydrolysis process from Figure 3.7 with an extra inhibitory effect due to ethanol presence in the medium. Cell biomass grows on glucose and xylose producing ethanol and CO_2. Furfural, 5-HMF and acetate inhibit sugar uptake, along with product inhibition by ethanol. Furfural, 5-HMF and acetate are also consumed by yeast leading to their depletion in time. All equations describing the reaction rates are based on [74] and reiterated in the supplementary material from Paper C.

pH and temperature are also key parameters in fermentation. The pH is disrupted due to CO_2 production, which leads to carbonic acid and bicarbonate production in the liquid phase. A pH calculator is necessary to predict the acidity of the medium, and to design a pH controller.

A local scheduler directs the flows according to the 4 phases, i.e. inoculum, fed-batch, batch, and unload. Another supervisory scheduler parallelizes multiple fermentation tanks, and ensures minimum interruptions of inflows and outflows from the overall fermentation process. Ideally these flows should be constant and continuous.

3.4.4 Fast pH Calculations with Guaranteed Accuracy for Dynamic Simulations

Enzymes and GMO yeast are very sensitive to pH variations following a Gaussian bell curve that describes their efficiency with a single optimal peak. A pH calculator is necessary for controlling and monitoring the enzymatic hydrolysis and fermentation processes, and is embedded into the modeling library. The calculator serves as a test bench for the pH controller designed in Chapter 5. The results from this section are not yet published and more implementation details are offered.

There are various ways of calculating the pH of a mixture, out of which the following three are the most used: (1) the alkalinity conservation approach, (2) the advancement method, and the charge balance approach [75]. The methods have several disadvantages: a numerical solver such as Newton-Raphson that often fails to converge (1), tracking a large number of states for each ion increasing model complexity (2), or a nonlinear equation solver that requires a long computational time (3) [75]. Such implementations turn pH calculations into bottlenecks in dynamic simulations.

The approach in this work is to use a reduced number of model states as in (1) and (3), and to avoid Newton-Raphson methods (where convergence can fail) and nonlinear equations solvers (which are slow). In contrast, the developed algorithm utilizes a modified bisection method for finding the root of the charge balance, which offers better convergence properties than the Newton-Raphson method and is faster than a solver of nonlinear equations. The charge balance is processed in order to express it as a function of only hydrogen ion concentration, which becomes the unknown. The bisection method is modified to advance in the pH logarithmic space while evaluating the charge balance in the state space. This method has the advantage that it always converges to a root with a preset accuracy if the function is monotonic and the solution can be bounded in an interval such that the function has opposite signs at its borders. Hence it has a guaranteed accuracy property under certain conditions. However, a higher accuracy increases the amount of convergence steps, which slows the computational time. The solution is to exploit the dynamic simulations and determine a tight interval around the root to be found at time k using the previous pH solution from time $k-1$. In dynamic simulations, the integration steps are relatively close to each other and the pH is not expected to change significantly in time, especially because it is a controlled variable kept to a reference point. The method has been found to be fast in reality, the pH calculations not appearing as bottlenecks in dynamic simulations anymore. Algorithm 1 illustrates the modified bisection method while Algorithm 2 shows how to compute the bounds around the sought solution.

Algorithm 1 Function for finding the pH of a mixture.

function PHCOMPUTE(S_T, ε, x_0) ▷ S_T is a vector with all species concentrations.
▷ ε is the solution tolerance.
▷ x_0 is the initial guess of the solution.
  $[l_b, u_b] \leftarrow$ COMPUTEBOUNDS(S_T, ε, x_0) ▷ Compute bounds around the solution.
  $\Delta x \leftarrow |l_b - u_b|$ ▷ Compute the width of the solution space.
  $n \leftarrow \lceil \log_2 \frac{\Delta x}{\varepsilon} \rceil$ ▷ Compute the necessary number of steps for convergence.
  $f_l \leftarrow pH(S_T, 10^{-l_b})$ ▷ pH value at the lower border.
  for $i \leftarrow 1, n$ **do**
    $x \leftarrow \frac{l_b + u_b}{2}$ ▷ Compute the center in logarithmic space.
    $\Delta x \leftarrow |l_b - u_b|$
    if $\Delta x \leq \varepsilon$ **then**
      break ▷ Solution within tolerance found.
    $f_x \leftarrow pH(S_T, 10^{-x})$ ▷ Evaluate pH in x.
    if $f_l \cdot f_x > 0$ **then**
      $l_b = x$ ▷ Change the left border.
      $f_l \leftarrow pH(S_T, 10^{-l_b})$ ▷ Update pH at the lower border.
    else
      $u_b = x$ ▷ Change the upper border.
  return x

Algorithm 2 Find bounds around pH initial guess.

function COMPUTEBOUNDS(S_T, ε, x_0)
  $l_b \leftarrow \max(1 \times 10^{-15}, x_0 - \varepsilon)$ ▷ Initialize lower bound l_b.
  $u_b \leftarrow \min(14, x_0 + \varepsilon)$ ▷ Initialize upper bound u_b.
  while $l_b > 1 \times 10^{-15} \vee U_B < 14$ **do**
    $f_l \leftarrow pH(S_T, 10^{-l_b})$ ▷ Evaluate pH in l_b.
    $f_u \leftarrow pH(S_T, 10^{-u_b})$ ▷ Evaluate pH in u_b.
    if $f_l \cdot f_u > 0$ **then**
      $\varepsilon \leftarrow \varepsilon \cdot 10$
      $l_b \leftarrow \max(1 \times 10^{-15}, x_0 - \varepsilon)$ ▷ Update lower bound l_b.
      $u_b \leftarrow \min(14, x_0 + \varepsilon)$ ▷ Update upper bound u_b.
    else
      break ▷ Bounds around solution were found.
  return l_b, u_b

The algorithm complexity is $O(\log n)$ given by the divide and conquer nature of the bisection method, where n is the size of the solution space. The worst case scenario corresponds to searching for the solution in the entire pH value set, i.e. $pH \in [0,14]$. This means that $\Delta x = 14$. If a two digits solution is sufficient then $\varepsilon = 0.01$. The number of converge steps then becomes $m = \lceil \log_2(1400) \rceil = 11$, meaning that a solution with accuracy ε is found in 11 steps. Such an accuracy is sufficient for controlling and monitoring the enzymatic hydrolysis and fermentation processes because a deviation of 0.01 from the optimal pH value has no impact on the process efficiency.

If the accuracy is increased, e.g. $\varepsilon = 1 \times 10^{-6}$, then the number of converge steps in the worst case scenario becomes $m = \lceil \log_2(14 \times 10^6) \rceil = 24$, which is still reasonable for nowadays computers. If the solution bounds are updated at each iteration using the previous solution as initial guess then the number of converge steps can be further reduced. For example, if the solution can be bounded within $\Delta x = 0.01$ units at each step then the solution space is significantly reduced and convergence takes place in $m = \lceil \log_2(1 \times 10^4) \rceil = 14$. In dynamic simulations with small integration steps the solution can be bounded even further following the method from Algorithm 2. The higher the accuracy is demanded the more performance can be offered by tightening the bounds around the solution.

3.4.5 Model Summary

The biorefinery model is complex and nonlinear including the most significant phenomena reported in the literature so far. Table 3.3 lists all features of the modeling library together with the publication that contains the full derivation details.

When interconnecting all library modules for describing the entire refinery, the model reaches a high complexity as in Table 3.4. The large number of states in pretreatment and enzymatic hydrolysis is due to CFD methods, which are used for discretizing big scale reactors. The inputs refer to feedstock mass flow, enthalpy, and composition, while the outputs comprise the outflow, enthalpy, and composition of the fermentation outstream. The table contains data for a single fermentation reactor but in reality there are several tanks running in parallel, which would increase the number of states in fermentation.

Table 3.3: Summary with all modeled phenomena.

Step	Features	Reference
Pretreatment	Steam atmosphere (single cell);	Paper D
	CFD for biomass layer;	Paper A
	Complex biomass conversion mechanism;	Paper A
	Sugars production: xylose, arabinose, glucose;	Paper A
	Pseudo-lignin production;	Paper A
	Xylooligomers intermediate product;	Paper A
	Organic acids production;	Paper A
	Inhibitors formation: furfural and 5-HMF.	Paper A
Enzymatic hydrolysis	Competitive conversion mechanism;	Paper B
	CFD for the first hydrolysis tank;	Paper B
	Enzymatic mixture parametrization;	Paper B
	Cellulose and hemicellulose hydrolysis;	Paper B
	Sugars production: glucose and xylose;	Paper B
	Cellobiose production;	Paper B
	Xylooligomers production;	Paper C
	Product inhibition;	Paper B
	Enzyme deactivation;	Paper C
	pH, temperature and severity dependence;	Paper C
	pH calculator;	Paper B
	Viscosity calculator.	Paper B
Fermentation	CSTR;	[74]
	Glucose and xylose uptake;	[74]
	Inhibitors uptake;	[74]
	Enzymatic hydrolysis;	[74]
	pH calculator.	[74]

Table 3.4: Model complexity summary: number of parameters, states, inputs and outputs. Half of the pretreatment outputs go to liquefaction, while the other half is connected to fermentation. The outputs from the enzymatic hydrolysis are connected to fermentation.

Model	Parameters	States	Inputs	Outputs
Pretreatment	17	298	10	36
Enzymatic Hydrolysis	46	257	18	19
Fermentation (1 tank)	33	25	37	25
Total	96	580	10	25

3.5 Model Analysis

This section summarizes the results from the sensitivity and uncertainty analysis for the pretreatment and enzymatic hydrolysis processes. The goal of the analysis is to identify modeling bottlenecks such as model parameters and feedstock components that contribute most to the accuracy of output predictions. Such an analysis offers support for design of experiments in order to properly estimate these parameters or identifies which feedstock components have to be measured accurately for reducing the uncertainty on estimates.

3.5.1 Sensitivity Analysis

Complex nonlinear models are most often over-parametrized but not all model parameters affect the outputs with the same weight. The sensitivity analysis is useful for process understanding and model reduction or simplification. Reducing the set of model parameters to an identifiable subset is a required step before performing parameter estimation based on real data.

3.5.1.1 Pretreatment

The sensitivity analysis results for the pretreatment process are shown in Figure 3.9. In reality a NIR instrument determines the composition of the pretreated slurry. The components of interest are cellulose, xylan, lignin, acetic acid, and furfural. The figure illustrates how each model parameter affects each of the measured components.

The method is a quantitative way of showing that the activation energies (and not the reaction constants) are the most sensitive model parameters but also ranks the parameters among themselves. Since all outputs are of interest for downstream, the cumulative indicator from the bottom right plot in Figure 3.9 is the most important result of this analysis. The parameter with the highest sensitivity is E_{Xo}, which is the activation energy for xylooligomers production due to hemicellulose hydrolysis. E_F and E_{Ac} follow indicating that furfural and acetic acid are important by-products of the pretreatment process. There is also some glucose production due to cellulose hydrolysis but much lower than hemicellulose hydrolysis. Pseudo-lignin and 5-HMF also appear with a lower impact on the outputs. The ranking shows that hemicellulose hydrolysis is the main phenomenon in the thermal reactor leading to furfural and acetic acid production as main by-products. Cellulose hydrolysis happens to a lesser extent because these fibers are protected by hemicellulose and lignin.

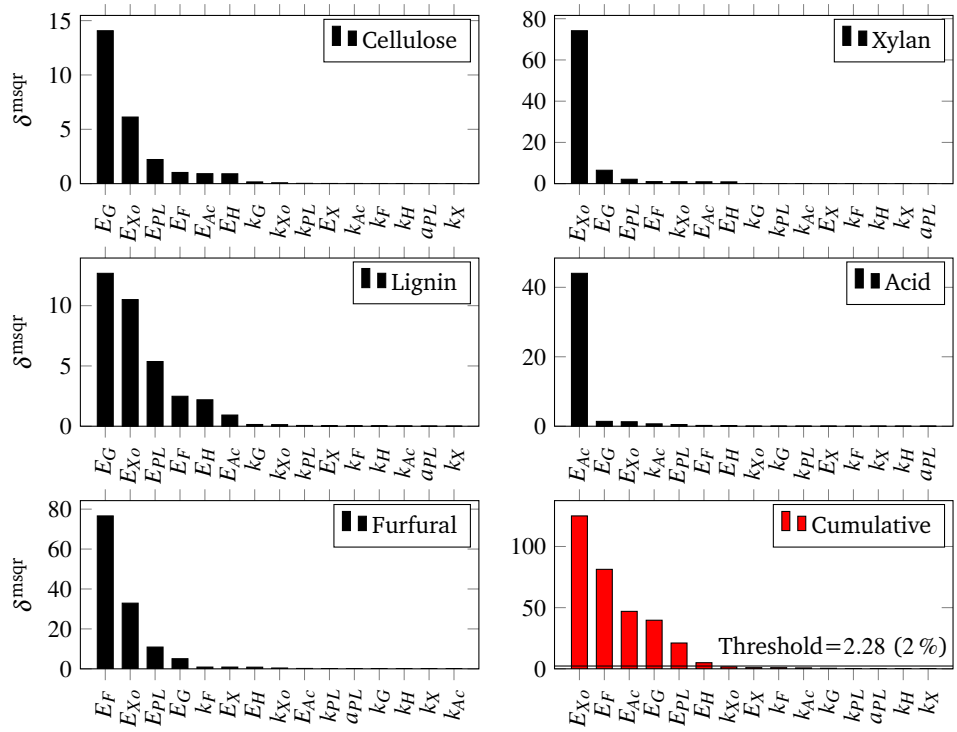


Figure 3.9: Sensitivity analysis by calculating the δ^{msqr} for each output with respect to each model parameter. Although the model has 15 parameters in total, only 6 parameters are of interest as indicated in the bottom right figure.

Out of the total count of model parameters, only 6 are found to be of interest. These parameters create an identifiable subset that is used for parameter estimation. The results of the estimation procedure together with the standard deviations and correlation matrix are shown in Paper A.

3.5.1.2 Enzymatic Hydrolysis

Following a similar approach, the sensitivity analysis for the enzymatic hydrolysis process is shown in Figure 3.10 where all parameters are ranked with respect to the cumulative sensitivity measure. The outputs of the liquefaction process are cellobiose, glucose and xylose concentrations, which are measured in reality with HPLC equipment based on grabbed samples every 6 h. The bottom subplot from Figure 3.10 is the most important one since all outputs are of interest. The first 4 most sensitive parameters are K_2, E_{M_1}, I_{X_2} and K_1, which refer to glucose

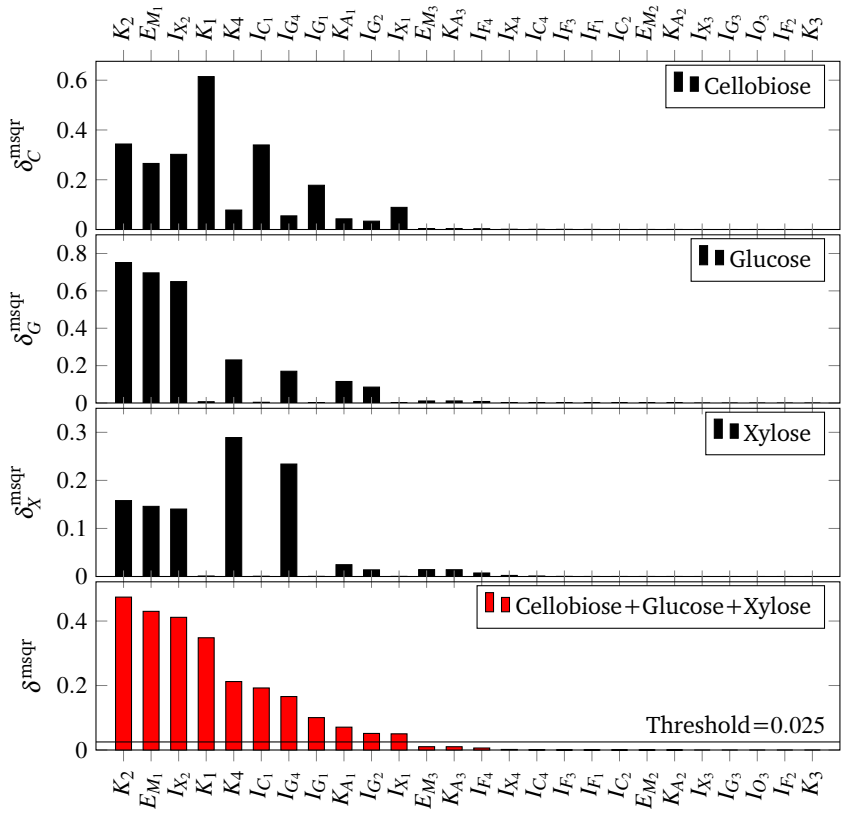


Figure 3.10: Sensitivity measure δ^{msqr} of model outputs with respect to kinetics parameters θ_K. The first 3 plots show δ^{msqr} for cellobiose (δ_C^{msqr}), glucose (δ_G^{msqr}) and xylose (δ_X^{msqr}). The bottom plot illustrates the overall sensitivity measure δ^{msqr} and a threshold that delimits the relevant parameters subset θ_S from the rest of the parameters.

production from cellulose, maximum adsorbed cellulase of endo-exo type, inhibition on glucose production by xylose, and cellobiose production reaction rate constant. The rest of the parameters have a lower sensitivity effect on outputs, and refer to xylose production and inhibition effects of cellobiose and glucose on the enzymatic hydrolysis.

The model parameter set is reduced to 11 parameters out of 46, a significant model reduction. Parameter estimation has not been attempted in this case due to the fact that the data are not persistent enough, and experimental design should be pursued first. Previous studies showed that such data lead to large confidence intervals on parameter estimates with high correlation [8].

3.5.2 Uncertainty Analysis

The uncertainty analysis is based on Monte Carlo simulations with LHS on model and feed parameters, which are treated in separated and combined scenarios. The results are grouped into pretreatment and liquefaction processes. The goal is to assess the model predictions' uncertainty and rank all sources of uncertainty in order to identify if feedstock components or kinetic parameters are most important for accurate estimates.

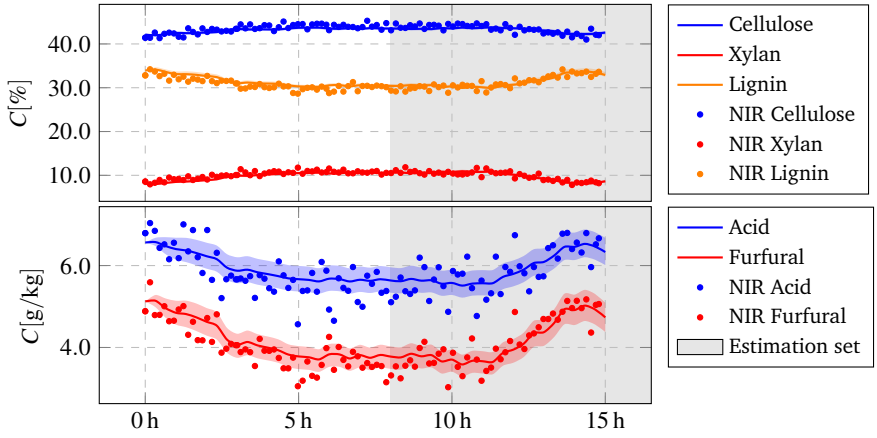
3.5.2.1 Pretreatment

Figure 3.11 illustrates the uncertainty analysis results for the pretreatment process. The solid bullets are NIR measurements performed on samples that were grabbed every 10 min during 15 h of operation. The thermal reactor temperature has been slightly changed, which caused different pretreatment conditions that modified the composition of pretreated fibers. The parameter estimation procedure uses the estimation subset identified with a grey area on the plots. The standard deviation and correlation matrix for model parameters is also obtained, which enables LHS with correlation control. All figures show the model fitting with 5th-95th percentiles uncertainty bounds for solids, i.e. cellulose, xylan, lignin (top plots), and solubles, i.e. acetic acid, furfural (bottom plots). The uncertainty bounds were calculated based on 200 Monte Carlo simulations. The solid line for each output represents the median.

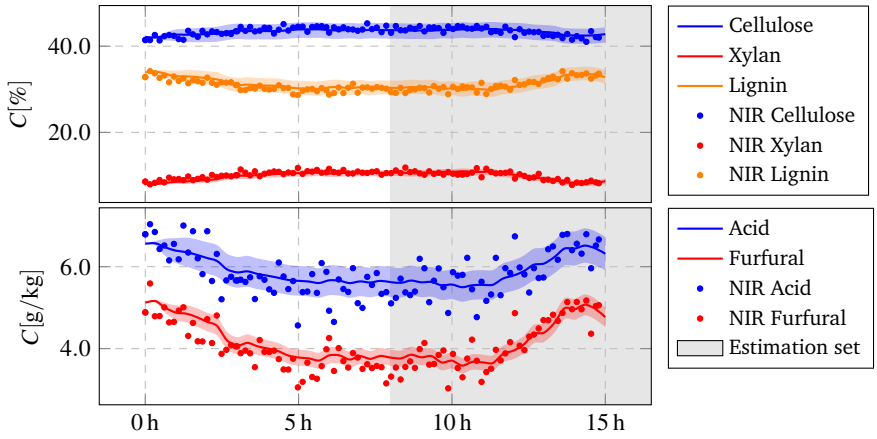
Figure 3.11(a) illustrates the uncertainty with respect to only model parameters. Figure 3.11(b) shows the model predictions when feed uncertainty is considered. The last figure 3.11(c) combines both sources of uncertainty, i.e. feed and model parameters. The results show that solid composition is more uncertain due to uncertainty in feed and not in model parameters because the uncertainty bounds from Figure 3.11(b) are larger than in Figure 3.11(a). Solubles uncertainty is relatively similar for both cases, with a reduced uncertainty in case of furfural due to feedstock. When combining both sources of uncertainty, the solubles prediction bounds increase capturing most of the measurements, while the solids prediction bounds remain almost the same as in the feed parameters case.

3.5.2.2 Enzymatic Hydrolysis

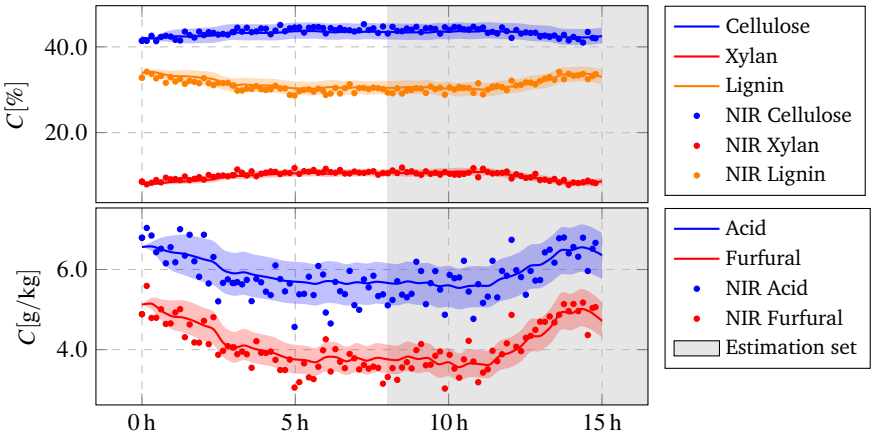
The uncertainty analysis for combined kinetics and feed parameters is displayed in Figure 3.12. Glucose, xylose and cellobiose were measured offline with HPLC equipment based on samples that were grabbed every approximately 6 h during 170 h of enzymatic hydrolysis operation. The process is continuous and, ideally, the sugar



(a) Final yield uncertainty with respect to model parameters.



(b) Final yield uncertainty with respect to feed parameters.



(c) Final yield uncertainty with respect to model and feed parameters (combined).

Figure 3.11: Pretreatment model uncertainty.

levels are constant lines in the absence of disturbances. However, the input fiber composition and the reactor hold-ups had variations throughout the data acquisition experiment. Changes in reactor levels lead to modifications in the retention time causing longer or shorter reaction times. These disturbances caused the variations seen in Figure 3.12.

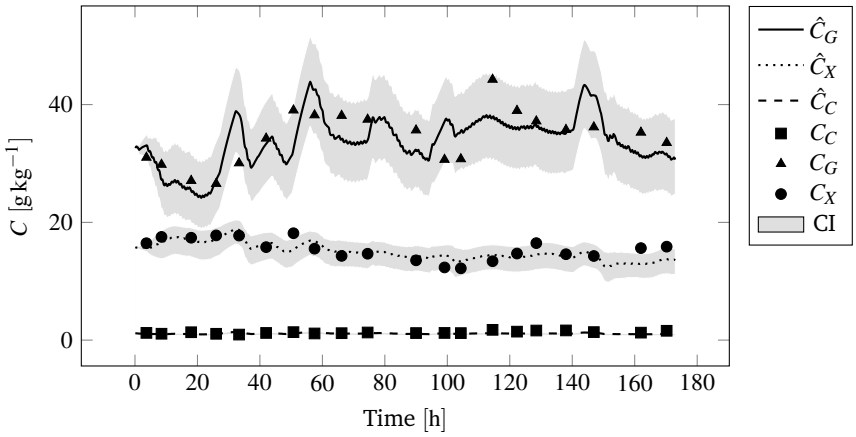


Figure 3.12: Cellobiose, glucose and xylose formation during the enzymatic hydrolysis process. $\hat{C}_C$, $\hat{G}_C$ and $\hat{X}_C$ are estimated concentrations of cellobiose, glucose and xylose as predicted by the model while C_C, C_G and C_X are sugar concentrations measured by the HPLC equipment. The gray area shows the 5th-95th percentile interval considering kinetics and feed uncertainties.

The model parameters' standard deviations and correlation matrix were obtained from [8], which enabled LHS with correlation control. The feed composition also varies due to NIR measurement errors, assumed to be around 5 % to 10 %. Another source of uncertainty refers to the fiber mass inflow, which was challenging to measure in reality. Most laboratory analysis results fit within the prediction bounds as shown in Figure 3.12. Cellobiose has a low level due to a high concentration of β-glucosidase. Xylose has a relatively accurate prediction level, while glucose has a larger confidentiality bound. More details regarding the generation of uncertainty data are included in Paper B.

3.6 Real-Time State Estimation of Biomass Pretreatment

A direct application of the pretreatment model is a state estimator, which has been published in Paper A. The motivation lies in compensating for modeling mismatches, in filtering the measurements provided by the NIR instrument, i.e. cellulose, xylan, lignin, acetic acid and furfural, and also to estimate unmeasured components in

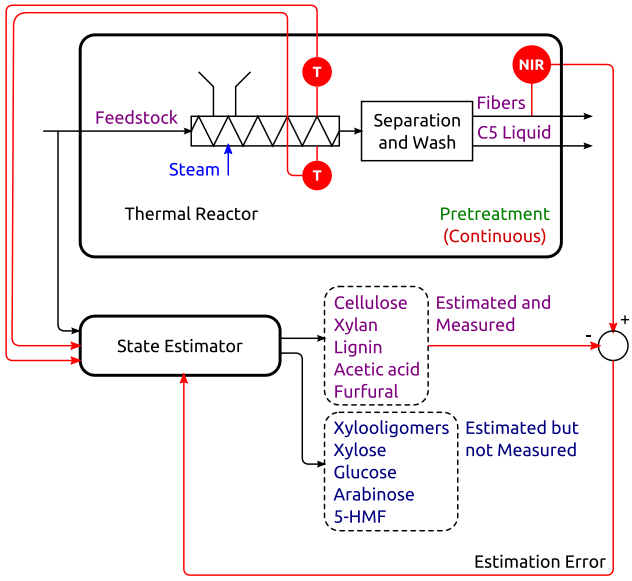


Figure 3.13: State estimator in biomass pretreatment.

pretreated biomass such as arabinose, xylose, xylooligomers, glucose, and 5-HMF. These components cannot be measured because either they cannot be detected by the NIR instrument or their concentration is too low as in the case of 5-HMF.

The block diagram of the observer is shown in Figure 3.13. Two sensors are placed at the top and bottom of the thermal reactor to measure the temperature of the steam and biomass layers. The temperatures are needed to construct the temperature gradient and to calculate the reaction rates in biomass conversion. The pretreated fibers are then measured by a NIR instrument every 10 min based on grabbed samples.

The state estimator has the structure of a high-gain observer embedding feedback from real measurements through a static Kalman gain. The estimation error is calculated for cellulose, xylan, lignin, acetic acid and furfural as the difference between the observer prediction and the NIR measurement. The entire methodology for calculating the static Kalman gain is included in Paper A.

The filtering and estimation results are shown in Figure 3.14. The top two plots illustrate the filtering of the NIR measurements, while the bottom plot displays the estimated content of unmeasured by-products. Xylose, xylooligomers and glucose are difficult to measure in reality in real-time. A soft sensor offers estimates of these components that can be used for monitoring the process in real-time while waiting for the HPLC analysis results of grabbed samples conducted off-line in the laboratory.

Other state estimators based on Kalman gains have been built for the temperature

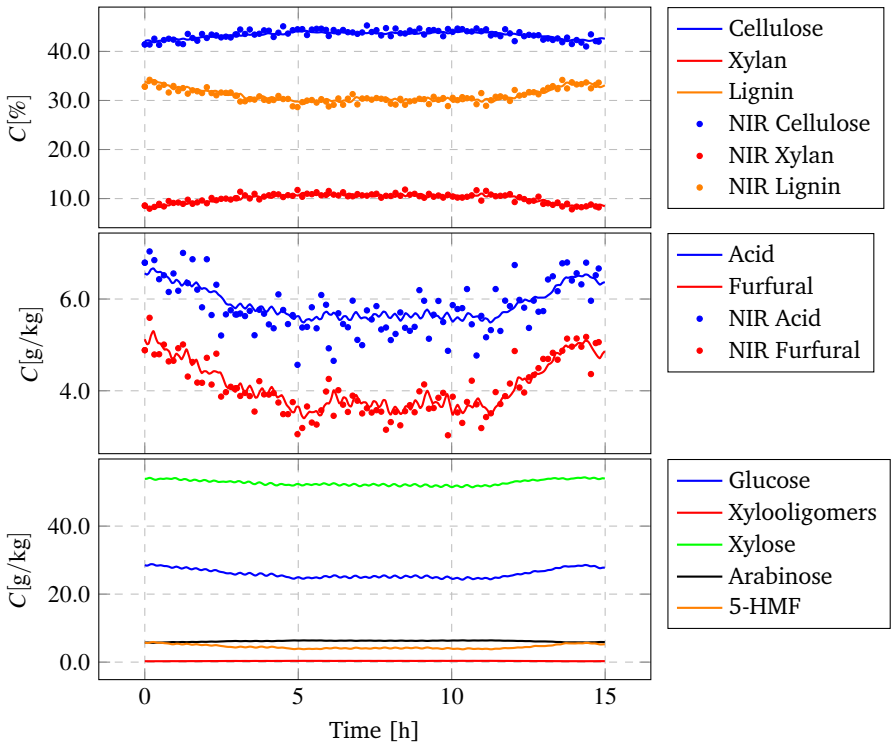


Figure 3.14: State estimator for biomass pretreatment.

gradients in the thermal reactor [71], for the enzymatic hydrolysis process [76], or fermentation [74]. Such observers are valuable for real implementation in order to filter measurements and to obtain estimates of all unmeasured biomass components.

3.7 Conclusions

This chapter presented the dynamic model library with focus on pretreatment and enzymatic hydrolysis. The co-fermentation model has been developed as a separate project included in [74] and reiterated in Paper C. The complex nonlinear mathematical models were then analyzed for sensitivity and uncertainty following a model analysis methodology that proved to be suitable for any complex industrial situation. The most important scientific novelty from this study was that the formulated models were validated against real data extracted from the Inbicon demonstration scale plant, showing the potential value of the dynamic library for optimization and control purposes at large scales. Also, the uncertainty analysis showed that feed parameters are a more important source of uncertainty than model parameters, and

focus should be placed on measuring feed composition more accurately.

The mathematical models need maintenance and should be extended as new phenomena appear in the scientific literature. A list with possible extensions include:

- Phenolics production in pretreatment and inhibition on the enzymatic activity [77];
- The irreversible deactivation of enzymes in time should be a function of temperature level and ethanol content [78], and not a constant;
- The fermentation temperature should actually follow a profile reflecting the best compromise between enzymatic activity and GMO yeast [79]. The calculation of the optimal temperature setpoint profile requires temperature dependent model parameters for fermentation [80], which are missing in this study.

Chapter 4

Process Optimization

4.1 Introduction

This chapter deals with designing a steady-state optimization layer for a large scale biorefinery. The objective is to maximize the operational profit by finding optimal key process parameters such as pretreatment temperature, enzyme dosage in liquefaction and yeast seed in fermentation. The refinery is treated in an integrated manner taking into account the trade-offs between the conversion steps. The detailed methodology and results are included in Paper C.

The method is compared against a traditional refinery operation, i.e. a fixed recipe with constant enzyme dosage and yeast seed that were determined based on a long time experience. Sensitivity and uncertainty analysis with respect to model and feed parameters increase the trust in the optimization layer. The chapter ends with conclusions and suggestions for future improvements.

4.2 Plantwide Optimization Methodology

Figure 4.1 illustrates the workflow interactions between the real plant and the layers for control, optimization and system identification. Each conversion stage from the real plant is equipped with specific sensors and actuators that enable fast feedback control for tracking the pretreatment temperature, enzyme dosage and yeast seed. An advanced reference tracking controller for biomass pretreatment temperature has been developed in [6]. An enzyme dosage regulator can be easily constructed with a flow meter and a variable valve mounted on the enzyme supply pipe. The enzymatic hydrolysis process is continuous and the enzyme dosage stabilizes to a non-zero value. The yeast seed controller requires a mass estimator and a control valve on the supply pipe. The fermentation process runs in batches,

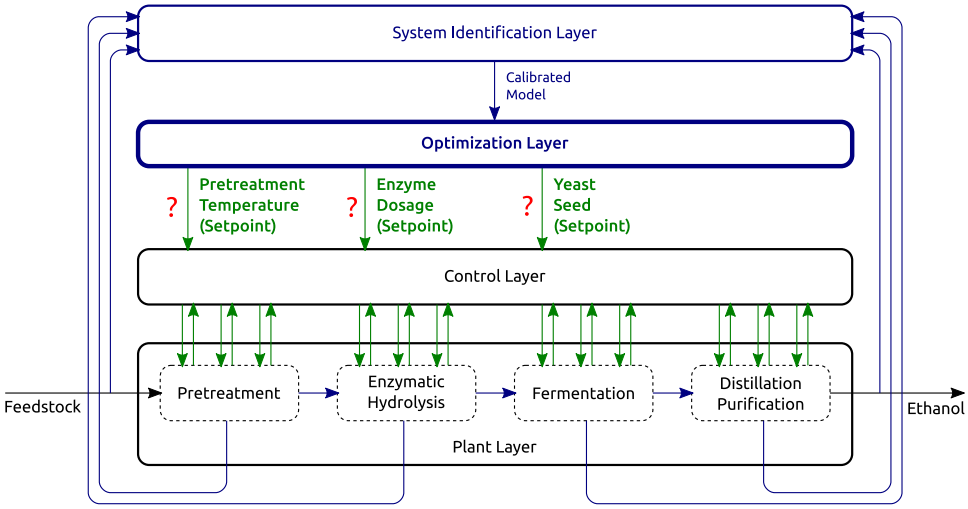


Figure 4.1: Block diagram showing the interaction between the optimization layer designed in this study and the real plant. The optimization layer calculates setpoints for pretreatment temperature, enzyme dosage and yeast seed. All models are calibrated by the system identification layer based on plant measurements.

and the yeast seed occurs only in the beginning of the inoculum phase until a certain mass of yeast has been accumulated. The control layer and the real plant interacts at a fast pace, i.e. in the order of seconds.

The control layer is designed for maximizing the biorefinery efficiency for a given constant throughput in order to support the optimization layer, which is common practice for a new plant [81]. The dynamic models from the previous chapter were validated against data collected at a throughput of $1\,\mathrm{t\,h^{-1}}$ of raw biomass. This flow rate represents 25 % of the refinery maximum capacity and was chosen in order to minimize the impact of pretreatment disturbances on fiber composition. At higher throughputs vertical temperature gradients appear in the thermal reactor that cause layers of different biomass composition [71]. The control layer also contains feedback loops that improve the conversion steps operation locally such as tank level, flow rate and temperature regulators, enzymatic and fermentation pH controllers, and an efficient scheduling control algorithm for fermentation.

The plantwide optimization methodology is inspired from [82], and applied to the biorefinery case from this project. A detailed description of the methods steps is given in Paper C, and summarized below:

1. Select the objective or cost function:

The cost function reflects the potential economical profit from 1 fermentation batch, which is defined as the difference between ethanol revenue and operating costs related to biomass, steam, enzymes, and yeast:

$$c(M_{Eth}, F_b, F_s, F_e, M_y) = M_{Eth}(t_f)P_{Eth} - (F_bP_b + F_sP_s + F_eP_e + M_yP_y) \tag{4.1}$$

The ethanol revenue per batch is calculated as mass of ethanol at the end of the batch phase $M_{Eth}(t_f)$ expressed in kg multiplied with its selling price P_{Eth} set in unitcost/kg. The refinery consumes the following utilities: yeast mass M_y (in kg), and F_b, F_s, and F_e (in $\mathrm{kg\,h^{-1}}$), which are the flow rates of feedstock, steam and enzymes, respectively. The price weights P_{Eth}, P_b, P_e, P_s, and P_y have a measuring unit such that the overall cost function c is expressed in a generic unitcost.

If the throughput is constant then the term F_bP_b is fixed and can be dropped from the cost function. Steam flow rate F_s reflects only pretreatment costs and does not include distillation. Although the purification costs are very high, they are related to water content, which is kept approximately constant in fermentation. After removing the fixed terms, the cost function is simplified to:

$$c(M_{Eth}, F_s, F_e, M_y) = M_{Eth}(t_f)P_{Eth} - (F_sP_s + F_eP_e + M_yP_y) \tag{4.2}$$

The numerical values of the weights are given in Table 4.1.

Table 4.1: Cost function weights (prices).

Parameter	Description	Value
P_{Eth}	Ethanol	5 unitcost/kg
P_e	Enzymes	25 unitcost/(kg/h)
P_s	Steam	1 unitcost/(kg/h)
P_y	Yeast	50 unitcost/kg

2. Identify the decision variables:

   Pretreated biomass composition is sensitive to the pretreatment temperature T_{tr} as shown in the sensitivity analysis from Paper A. The liquefaction and fermentation processes are dependent on enzymes flow and yeast seed, which have high acquisition price and should be used in an efficient way. These decision variables are placed in a vector form:

$$u = \begin{bmatrix} T_{tr} & F_e & M_y \end{bmatrix}^\top \tag{4.3}$$

3. Process model constraints, and bounds for decision variables:

   The continuous processes of pretreatment and enzymatic hydrolysis reach steady states, which are calculated by solving:

$$\dot{x} = f(x,u) \quad \text{when} \quad \dot{x} = 0 \tag{4.4}$$

   Function $f(x,u)$ describes the time dynamics of these processes as expressions of state vector x and input u.

   The fermentation model is a batch process and only the final states are of interest:

$$x_f(t_f) = \int_0^{t_f} h(x_f, u_f)dt \tag{4.5}$$

   where $h(x_f, u_f)$ is the fermentation dynamic model as function of states x_f, and inputs u_f. Time constant $t_f = 190\,\text{h}$ represents the end of the batch phase.

   The decision variables are bounded as follows:

$$\begin{array}{rcccl} 150 & \leq & T_{tr} & \leq & 210\,^{\circ}\text{C} \\ 10 & \leq & F_e & \leq & 1000\,\text{kg}\,\text{h}^{-1} \\ 10 & \leq & M_y & \leq & 1000\,\text{kg} \end{array} \tag{4.6}$$

   reflecting a wide range of operation.

4. Formulate and solve the overall optimization problem:

$$\begin{array}{rcccl} \max\limits_{T_{tr},F_e,M_y} & & M_{Eth}(t_f)P_{Eth} - (F_sP_s + F_eP_e + M_yP_y) & & \\ \text{subject to} & & 0 = f(x(t),u(t)) & & \\ & & \dot{x}_f = h(x_f,u_f) & & \\ 150 & \leq & T_{tr} & \leq & 210\,^{\circ}\text{C} \\ 10 & \leq & F_e & \leq & 1000\,\text{kg}\,\text{h}^{-1} \\ 10 & \leq & M_y & \leq & 1000\,\text{kg} \end{array} \tag{4.7}$$

   The following solution is found:

$$z_o = \begin{bmatrix} T_{tr} \\ F_e \\ M_y \end{bmatrix} = \begin{bmatrix} 172\,^{\circ}\text{C} \\ 110\,\text{kg}\,\text{h}^{-1} \\ 142\,\text{kg} \end{bmatrix} \tag{4.8}$$

   with a cost value of:

$$c_o = 76\,714\,\text{unitprofit} \tag{4.9}$$

   The cost value does not include feedstock and distillation costs.

5. The sensitivity and uncertainty analysis is treated in the next section. This analysis iterates through a range of pretreatment temperatures in order to gain process insight into how pretreatment conditions affect the costs, profit and optimal solution of the optimization problem.

4.3 Sensitivity and Uncertainty Analysis

The model is over-parametrized with 96 parameters. The goal of the sensitivity analysis is to rank all model parameters with respect to their sensitivity on the profit curve at the optimal solution, and subtract a reduced subset in order to identify the dominant phenomena and conversion steps. The scope of the uncertainty analysis is to find confidence intervals for costs, profit and optimal solution considering sources of uncertainty for both model and feed parameters. The uncertainty analysis is carried for a wide range of pretreatment conditions in order to capture and understand the trade-offs between the conversion steps.

The methodology for the sensitivity and uncertainty analysis is shown in Algorithm 3. The steps are similar to the ones from the model analysis case in Chapter 3. The model parameters are first initialized and then the optimal solution of problem (4.7) is found. Next, the sensitivity of the cost function with respect to all model parameters is calculated. This analysis uses a non-dimensional sensitivity measure for steady state signals defined as [8, 64]:

$$\delta_k = \frac{\partial c}{\partial \theta_k} \frac{\theta_k}{c_{ss_k}} \tag{4.10}$$

where c is the cost function, θ_k is the kth model parameter, and c_{ss_k} is the stabilized value of the cost function. Model parameters are ranked with respect to δ_k, and a subset of significant parameters θ_R is determined.

The uncertainty analysis is carried for both model and feed parameters, in separated and combined scenarios. The model parameters are sampled following LHS with correlation control, while feedstock composition is uniformly sampled covering a 5 % uncertainty range. The standard deviations and correlation matrices for model parameters are obtained from published data [5, 8]. Algorithm 4 then runs for each set of model and feed parameters in order to determine the costs, profit, and optimal solution as functions of pretreatment temperature. The uncertainty bounds are calculated based on the Monte Carlo simulations.

Algorithm 4 iterates through pretreatment temperatures within the range 160 °C to 180 °C with a 1 °C resolution. For each temperature, another optimization problem is formulated where pretreatment is decoupled and liquefaction and fermentation processes are treated in an integrated manner. This new optimization problem is

Algorithm 3 Sensitivity and uncertainty analysis with combined model and feed uncertainty.

1: Initialize model parameters θ and feed composition C_b.
2: Optimal deterministic solution: $[T_{tr}, F_e, M_y] \leftarrow$ (4.7).
3: Sensitivity analysis of the cost function in $[T_{tr}, F_e, M_y]$: calculate δ_k as in equation (4.10).
4: Rank all parameters with respect to δ_k.
5: Select a subset θ_R such that θ_R is above a threshold.
6: Set standard deviations and correlation matrices for θ_R.
7: $\theta \leftarrow$ LHS of θ_R with correlation control to generate N sets of model parameters.
8: Set bounds for feedstock composition.
9: $C_b \leftarrow$ Uniform LHS for feedstock composition to generate N sets of compositions.
10: **for** Each set of model and feed parameters **do**
11: $\left[z_i, c_{p_i}, c_{eh_i}, c_{f_i}, c_i\right] = \text{COSTS}(\theta_i, C_{b_i})$
12: Calculate the 5th, median and 95th percentiles for profit, costs, and optimal solution.

Algorithm 4 Calculate optimal costs and profit with respect to pretreatment temperature given a fixed set of model parameters θ and feedstock composition C_b.

1: **function** $\left[z, c_p, c_{eh}, c_f, c\right] = \text{COSTS}(\theta, C_b)$
2: Set a range of pretreatment temperatures $T_{tr} \leftarrow 160\,^{\circ}\text{C} : 1\,^{\circ}\text{C} : 180\,^{\circ}\text{C}$
3: Set initial solution guess to $z_0 \leftarrow [100\,\text{kg}\,\text{h}^{-1} \quad 80\,\text{kg}]^{\top}$
4: **for** Each temperature in T_{tr} **do**
5: Run pretreatment process at temperature T_{tr_i} and obtain composition of pretreated fibers and C5 liquid, and steam inflow into the thermal reactor F_{s_i}.
6: $z_i \leftarrow$ Solution of optimization problem (4.11) given the pretreated fibers composition and C5 liquid from previous step as inputs. Use as initial guess the solution from previous iteration z_{i-1}.
7: Calculate mass of ethanol at final fermentation time: $M_{Eth_i} \leftarrow M(t_f) \cdot C_{Eth}(t_f)$ where t_f is the final batch time, $M(t_f)$ is the reactor mass in kg at time t_f, and $C_{Eth}(t_f)$ is the ethanol concentration at time t_f in $\text{g}\,\text{kg}^{-1}$.
8: Enzyme dosage: $F_{e_i} \leftarrow z_i(1)$.
9: Yeast seed: $M_{y_i} \leftarrow z_i(2)$.
10: Calculate pretreatment cost: $c_{p_i} \leftarrow F_{s_i} P_s$.
11: Calculate liquefaction cost: $c_{eh_i} \leftarrow F_{e_i} P_e$.
12: Calculate fermentation cost: $c_{f_i} \leftarrow M_{y_i} P_y$.
13: Calculate revenue: $r_i \leftarrow M_{Eth_i} P_{Eth}$.
14: Calculate profit $c_i \leftarrow r_i - (c_{p_i} + c_{eh_i} + c_{f_i})$.

similar to (4.7) but without T_{tr} from the decision variables vector u, and $F_s P_s$ from the cost function:

$$\begin{aligned} \max_{F_e, M_y} \quad & M_{Eth} P_{Eth} - (F_e P_e + M_y P_y) \\ \text{subject to} \quad & 0 = f(x(t), u(t)) \\ & \dot{x}_f = h(x_f, u_f) \\ 10 \quad \leq \quad & F_e \quad \leq \quad 1000\,\mathrm{kg\,h^{-1}} \\ 10 \quad \leq \quad & M_y \quad \leq \quad 1000\,\mathrm{kg} \end{aligned} \tag{4.11}$$

The sensitivity analysis results are shown in Figure 4.2. The model parameters count is reduced to 22 significant parameters from 97 in total. Pretreatment and fermentation parameters are the most important ones: E_F related to furfural formation in pretreatment, which inhibits ethanol production, Y_{PS_G} or the ethanol yield per glucose, E_G or cellulose hydrolysis in pretreatment, $q_{Ac_{Max}}$ related to acetate uptake, which also inhibits fermentation, and Y_{PS_X} and Y_{XS_G} representing ethanol production from xylose, and biomass growth on glucose. Liquefaction parameters have a lower sensitivity, the first ones being R_B and K_2, i.e. enzymatic activity dependence on pretreatment severity, and glucose production due to cellulose hydrolysis.

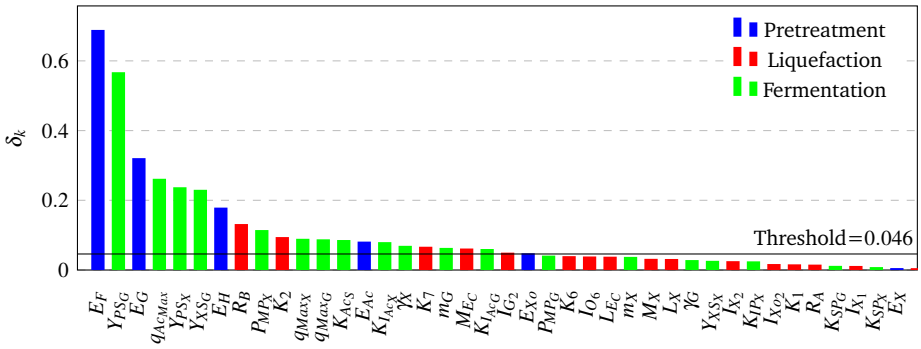
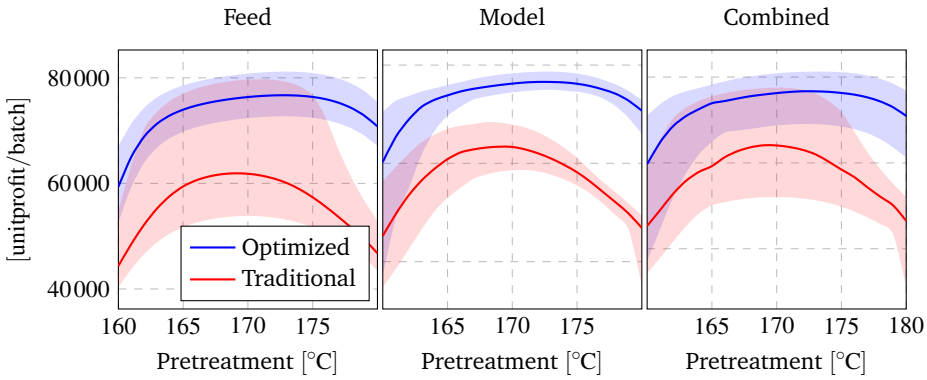
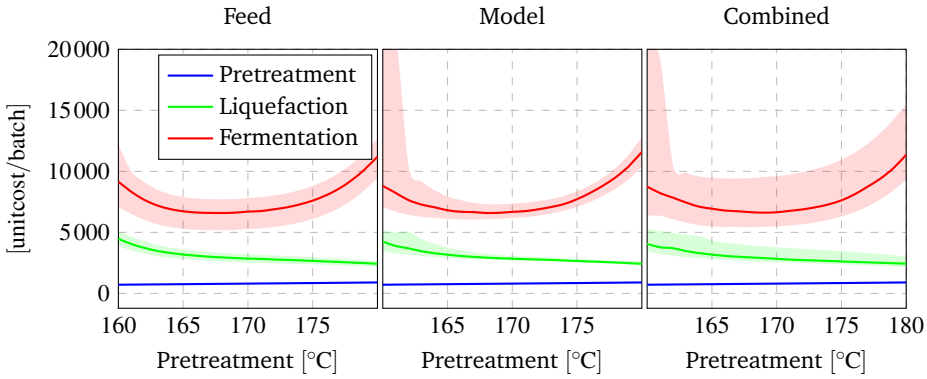


Figure 4.2: Sensitivity measure δ_k of profit value with respect to model parameters.

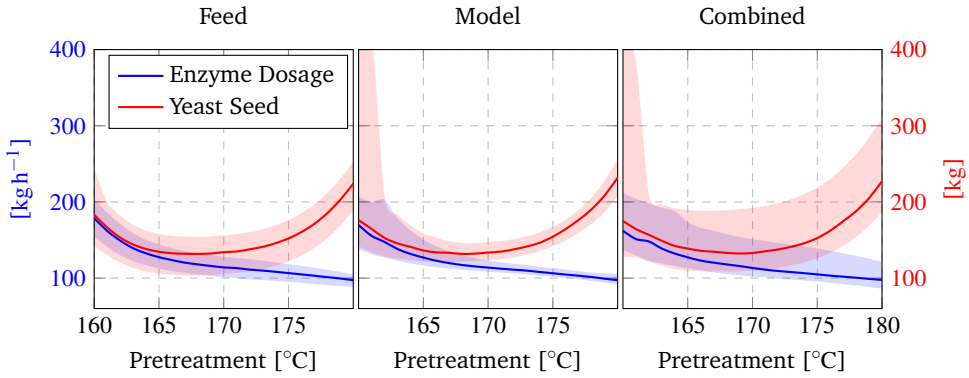
The uncertainty analysis results are presented in Figure 4.3. The potential profit for one fermentation batch excluding feedstock and distillation costs is shown in Figure 4.3(a). The optimized curve is in fact the value of the cost function at the optimal solution when treating the refinery in an integrated manner, while the traditional curve represents the profit following an operation with a fixed recipe for pretreatment, hydrolysis and fermentation without any adjustments. The traditional operation is usually established when trying to optimize the refinery by separating the conversion steps. The optimized curve offers an overall improvement of about 18 % over the traditional operation. The profit is reduced at low temperatures



(a) Potential profit with uncertainty for one fermentation batch.



(b) Refinery costs for each biomass conversion step.



(c) Solution of the optimization problem.

Figure 4.3: Optimal costs, potential profit, and solution of the optimization problem.

because the biomatrix is not sufficiently open, and hemicellulose is not hydrolyzed enough. A low pretreatment temperature requires more enzymes and yeast to compensate for hardly accessible cellulose, also indicated in the solution plots from Figure 4.3(c). At higher pretreatment temperatures inhibitors become significant and decrease ethanol yield causing a lower profit. The optimal pretreatment temperature is found to be around 172 °C. Another important aspect is that feed dominates model parameters uncertainty. Accurate measurements on feedstock are required to decrease the uncertainty on the potential profit of the refinery process. Model uncertainty also decreases as temperature increases due to the fact that biomatrix opening reduces its significance on model predictions. Overall the optimized profit curve has a lower uncertainty than the traditional operation because the enzyme dosage and yeast seed are adjusted with respect to pretreatment conditions.

Figure 4.3(c) shows the solution of the optimization problem as a function of pretreatment temperature. The amount of enzymes decrease as pretreatment temperature increases showing that a high pretreatment is beneficial for enzymatic hydrolysis, while negative for fermentation due to inhibitors creation. The yeast curve has a U shape. At low temperatures, a higher amount of yeast accelerates the simultaneous saccharification and fermentation process while at higher temperature more yeast compensates for the inhibitors. The refinery costs are proportional to the solution of the optimization problem, and detailed in Figure 4.3(b).

4.4 Conclusions

This chapter presented the results from Paper C, i.e. a study on economical optimization of a large scale second generation biorefinery in a simulated environment. The optimization procedure makes use of steady-state models for pretreatment and enzymatic hydrolysis, and a dynamic fermentation model. The process key parameters are the pretreatment temperature, enzyme dosage in liquefaction, and yeast seed in fermentation. A high pretreatment temperature was found to be positive for the performance of enzymatic hydrolysis but negative for ethanol yield due to creation of inhibitors. Uncertainties in kinetics of pretreatment, liquefaction and fermentation were quantified as negligible on the economic objective function around the optimal operational point. The main source of uncertainty was found to be in the inflow feed composition.

The enzymatic hydrolysis and fermentation are connected in series in this study as in the second version of Inbicon. However, the same methodology can be applied in the case of a fermentation broth recycle that would enable enzymes reuse [83].

The optimization was carried at constant given throughput. As a future improve-

ment, the biorefinery throughput can be added as a degree of freedom that would allow searching for a higher throughput such that profit is maximized [84]. In reality, the optimal profit value is slightly degraded due to the compromise between hydrolysis and optimal fermentation temperature [85], which decreases ethanol yield or would require a higher amount of enzymes and yeast to compensate. An optimal temperature profile can be calculated for maximizing the ethanol yield [79]. In fermentation, feed rate control boosts the performance by maximizing the specific growth rate while avoiding accumulation of inhibitors [86]. Also, in the case of glucose fermentation, the amount of necessary yeast can be significantly reduced by calculating an optimal feed rate profile during the fed-batch phase while avoiding accumulation of inhibitors as in Section 5.4.

For ensuring the successfulness of the optimization method, the control layer and the process architectural setup should be solved in an integrated manner and not sequentially as is common practice in the industry [87]. The contribution of this study is to offer support for additional changes that are necessary to improve the biorefinery design. The results from this chapter were based on several assumptions regarding available measurements and control strategies, which should be included in future plant versions: feedstock NIR measurements, and control on enzymes flow rate and yeast dosage.

Chapter 5

Advanced Process Control

5.1 Introduction

This chapter presents the main results from two IEEE peer-reviewed conference publications included in the appendix as Papers E and D, as well as unpublished results regarding an optimal feed rate profile of liquefied fibers in glucose fermentation that minimizes the yeast seed and avoids accumulation of inhibitors. The IEEE conference contributions deal with two applications of the L1 adaptive output feedback controller: one for regulating the temperature in the pretreatment thermal reactor, and another one for keeping the pH level close to its optimum in the enzymatic hydrolysis process.

5.2 Pretreatment Temperature Control

The sensitivity analysis for the hydrothermal biomass pretreatment model from Section 3.5.1.1 showed that the activation energies of conversion reactions in the thermal reactor are sensitive for pretreated biomass composition. That is also the reason why the pretreatment temperature is one of the decision variables in the optimization layer. These facts show that a temperature controller is necessary to ensure little variations from the desired reaction temperature, and to transit smoothly to the setpoint given by the optimization layer.

The particle pump pressurizes biomass till the thermal reactor pressure based on the following cycle: an amount of soaked fibers is first trapped in the particle pump chamber with the help of a pushing snail; the thermal reactor pressure is then reached by opening a steam valve from a pipe that interconnects the reactor and the particle pump; once the same pressure level is reached, the steam valve is closed, and the biomass is released inside the thermal reactor. The total duration of the

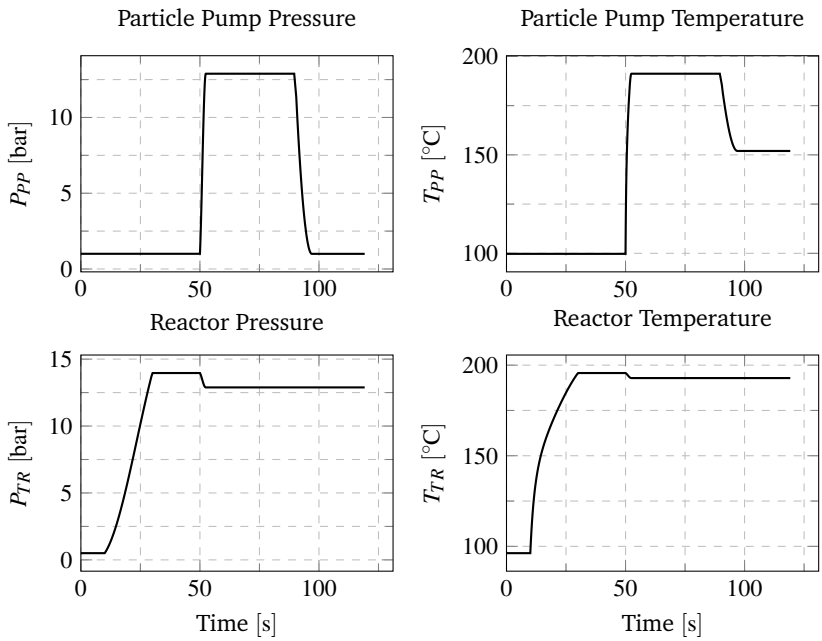


Figure 5.1: Open loop simulation. The top plots show the pressure and temperature in the particle pump while the bottom plots display the pressure and temperature in the thermal reactor.

cycle is of approximately 2 min. Pressurizing the particle pump with recycled steam disturbs the pressure in the thermal reactor causing the reaction temperature to drop as a consequence. Variations in reaction temperature lead to irregularities in the composition of the pretreated biomass [71].

Figure 5.1 illustrates the open loop operation of the particle pump and thermal reactor tandem. The particle pump pressure P_{PP} and temperature T_{PP} are plotted in the top graphs while the reactor pressure P_{TR} and temperature T_{TR} are shown in the bottom plots. The thermal reactor is pressurized till 13 bar with fresh steam starting at 10 s. The steam valve connecting the thermal reactor to the particle pump is opened at 50 s, which causes the reactor pressure and temperature to drop. At 90 s the particle pump opens the evacuation valve and the pressure is restored to the atmospheric level.

The scope of this part of the study is to design a controller that rejects the steam recycle disturbances, ideally keeping the reactor temperature constant. Nonlinear valve characteristics are considered due to the fact that the pressure drop across the steam valve varies significantly, i.e. from 0 bar to 13 bar. The controller has to be operational in multiple nominal points depending on the reactor load, which

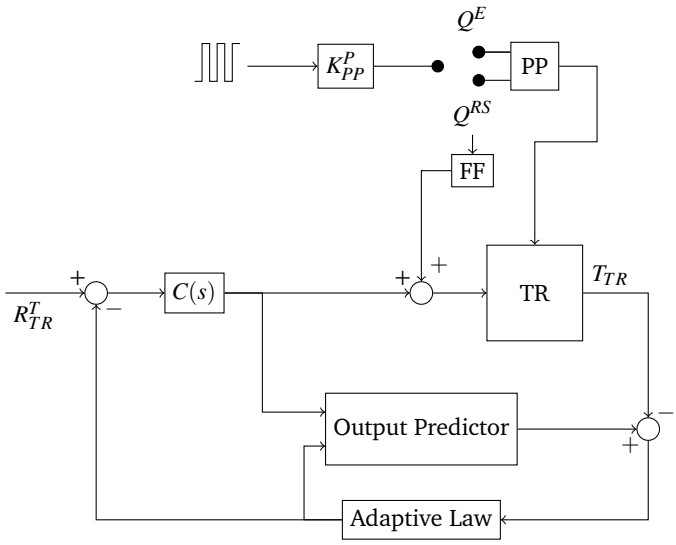


Figure 5.2: The control block diagram: L1 adaptive output feedback controller connected to a thermal reactor and particle pump tandem.

influences the steam expansion volume. The L1 adaptive controller appeared as a novel control strategy in 2007 ensuring fast adaptation with guaranteed robustness [88]. Due to the fact that only a temperature sensor is desired to be used for control purposes, an output feedback structure [10] is selected as in Figure 5.2.

The upper part of Figure 5.2 shows the particle pump (PP) switch controller K^P_{PP}, which does not require feedback because the pressure in both the reactor and particle pump eventually equalizes when opening the recycle steam valve that interconnects the two units. The operation sequence for the recycle and evacuation valves resembles a pulse signal. The recycle steam flow rate is measured as Q^{RS}, and a feed-forward term (FF) is calculated for finding the corresponding stroke of the fresh steam valve to compensate for the extracted steam. The openings of the recycle and fresh steam valves are not identical because the pressure drop across them is not the same.

The block diagram of the L1 controller is similar to the one of a Model Reference Adaptive Controller (MRAC) with the addition of a filter $C(s)$ on the control signal. The output predictor contains the desired dynamics in closed loop, while the adaptive law is set to be fast. The role of filter $C(s)$ is to improve the stability margins and partially remove the chattering introduced by a fast adaptation law. Hence the L1 adaptive controller offers robustness and fast adaptation. The main challenge becomes in tuning a proper and stable filter. The L1 norm is used to prove that there

exists uniform bounds for system states and control signals [10].

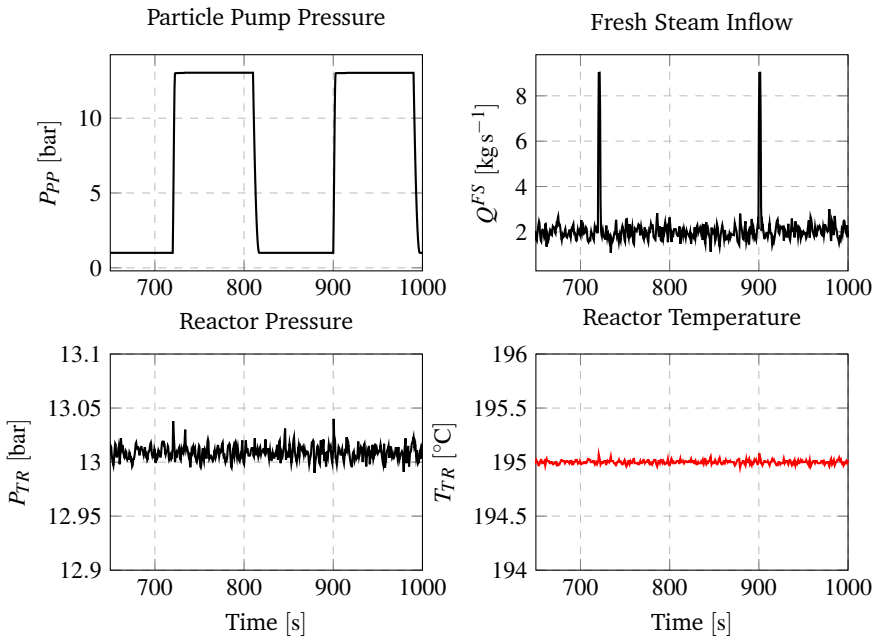
In the context of the pretreatment temperature application, the desired closed loop dynamics are set to a first order system parametrized in time constant $1/m$, $C(s)$ is a first order low-pass filter with time constant $1/c$, and the adaptation gain is set to Γ. The tuning approach in this paper is to formulate an optimization problem that minimizes the IAE with respect to the controller parameters, i.e. m, c and Γ:

$$\begin{aligned} \min_{m,c,\Gamma} \quad & \int_0^\infty |r(t) - y(t)| \\ \text{subject to} \quad & \dot{x} = f(x(t), u(t)) \\ & y = g(x(t)) \end{aligned} \tag{5.1}$$

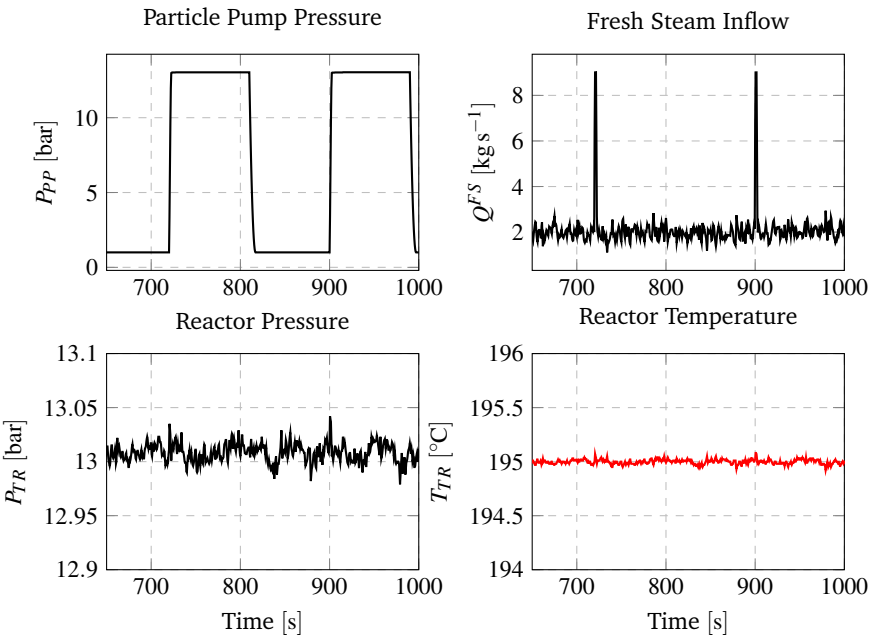
The system dynamics are calculated by integrating $\dot{x} = f(x(t), u(t))$ in time where x is the states vector and u is the fresh steam valve opening or the manipulated variable (MV). The controlled variable (CV) is the reactor temperature $y = g(x(t))$. The overall process is a Single Input Single Output (SISO) system. The problem is solved numerically by evaluating the cost function in simulation scenarios that correspond to a normal operation cycle, i.e. the particle pump extracting steam from the thermal reactor, and the controller restoring the pressure in the reactor.

After tuning the controller, the performance of the closed loop system is evaluated in three scenarios: (1) normal operation with non-zero mean white noise for condensation, (2) normal operation with zero mean measurement noise, and (3) reference tracking with condensation noise. The results of the disturbance rejection scenarios, i.e. (1) and (2), are shown in Figures 5.3(a) and 5.3(b) where two pressurization cycles are emphasized. The reactor temperature appears undisturbed in all scenarios around the value of 195 °C even when the particle pump extracts steam from the thermal reactor, i.e. at timestamps 720 s and 900 s. The pressure and temperature have little deviations from their nominal values, which ensures a uniform pretreatment environment. The fresh steam inflow is displayed in the top right plots. The flow spikes correspond to the time moment when the recycle steam valve opens for pressurizing the particle pump. The steam inflow is non-zero in idle operation due to steam condensation inside the thermal reactor.

The reference tracking scenario from Figure 5.4 includes steps of 5 °C and 10 °C around the nominal operation of 195 °C. The first reference change is at 200 s to 200 °C, followed by a −10 °C change until 190 °C at 400 s. The reference returns to the nominal value at 600 s. The closed loop dynamics resemble a first order system as imposed by the model reference adaptive controller. The positive increments in temperature are achieved by injecting more steam into the reactor while the negative step is obtained by closing the fresh steam valve and letting condensation decrease the pressure.



(a) Normal operation cycle with condensation noise.



(b) Normal operation cycle with measurement noise.

Figure 5.3: Disturbance rejection scenarios.

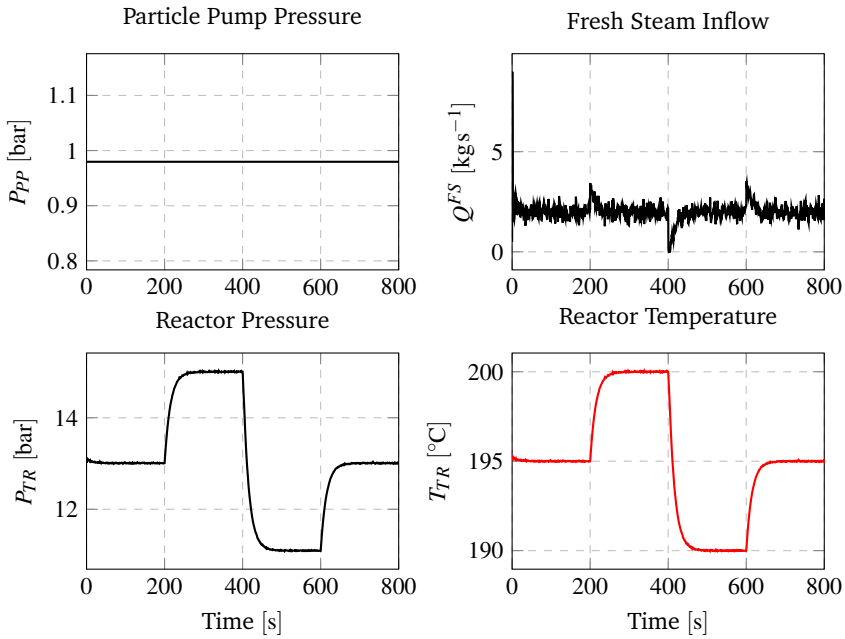


Figure 5.4: Reference tracking with condensation noise.

The closed loop system performs satisfactory both in disturbance rejection by reducing the reactor temperature variations, and in reference tracking ensuring a fast change to a new temperature level. The controller setpoint is given by the optimization layer designed in Chapter 4. More details about the adaptive controller design including closed loop transfer function analysis can be read in Paper D.

5.3 Enzymatic pH Control

pH is another process key parameter because enzymes and yeast are highly sensitive to the pH level following a Gaussian bell curve. A pH controller is requested to keep the pH of the mixture at the optimal level, e.g. the peak of the enzymatic activity curve. The enzymatic hydrolysis pH setpoint is taken from the enzymes manufacturers, and remains constant throughout the entire process. In simultaneous saccharification and fermentation an optimal pH profile exists as a trade-off between enzymes and yeast optimal pH [89]. In this case, pH reference tracking is of interest.

The pH is disturbed by organic acids that are produced during pretreatment due to hemicellulose removal, and also during enzymatic hydrolysis of the remaining hemicellulose in the liquefaction tanks. In fermentation the pH is disturbed due to CO_2 production, and organic acids formation during saccharification of the

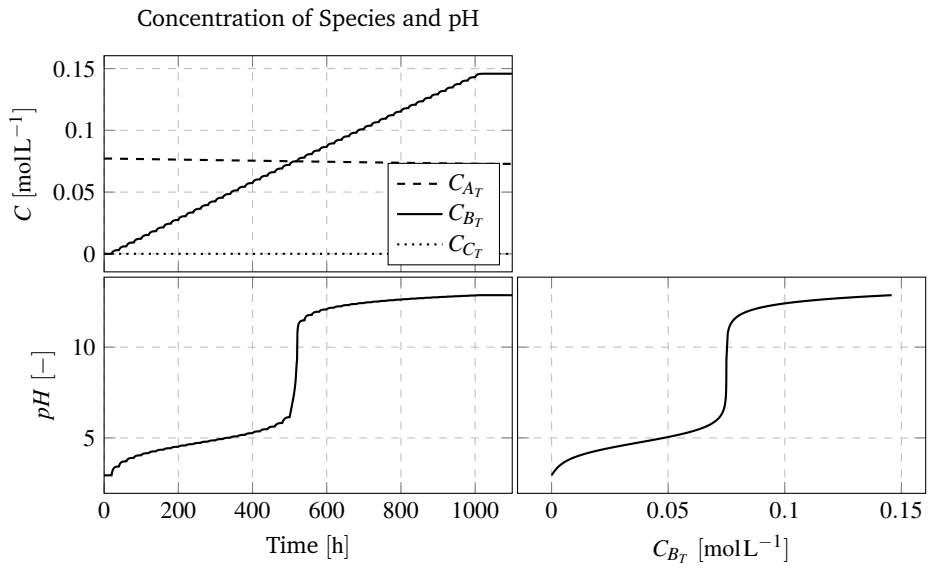


Figure 5.5: Titration simulation. The main challenge lies in the nonlinearity of the titration curve shown in the bottom right plot.

remaining hemicellulose. This section presents the results from Paper E where an L1 output feedback pH controller is implemented for the enzymatic hydrolysis process. The same control methodology can be applied to develop a pH controller for the fermentation process.

Enzymes are sensitive to the pH of the mixture following a bell shaped dependency curve [11]. The control objective is to keep the pH level close to its optimum, which is set according to the manufacturer's advice, e.g. $pH = 5 \pm 0.1$. The pH is regulated by pumping a strong base, e.g. NaOH, from a storage tank based on a flow rate setpoint given by the adaptive controller developed in this section. The pump flow rate control loop is considered trivial and neglected in this work.

The main challenge in reaching the pH control objective is the nonlinearity of the titration curve. Figure 5.5 illustrates a simulated titration experiment where base is added in small steps to the liquefaction reactor. The top left plot displays the concentrations of organic acid C_{A_T}, base C_{B_T}, and carbonic species C_{C_T}, all in $\mathrm{mol\,L^{-1}}$. Concentration of base linearly increases with equal increments while acid and carbonic species remain constant. The bottom left plot shows the pH value as a function of time as the simulated experiment progresses while the bottom right graph draws the titration curve or the pH as a function of base concentration C_{B_T}.

The block diagram of the closed loop system is illustrated in Figure 5.6. The pH is measured directly, and used to build the feedback action. A feed-forward term

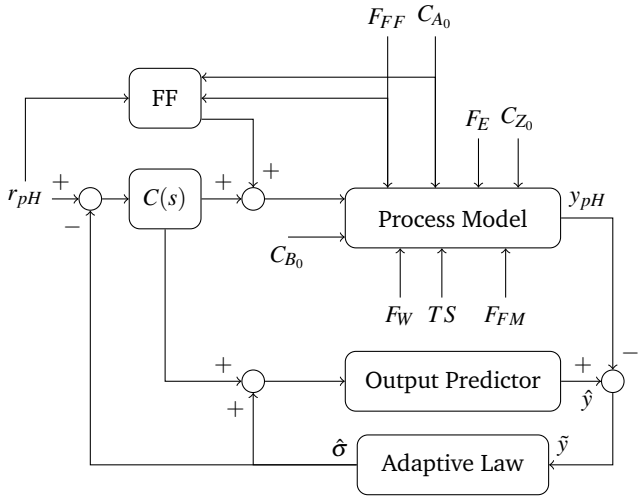


Figure 5.6: Closed loop system with an L1 adaptive output feedback controller.

is constructed knowing the inflow rate of fibers F_{FF}, the total solids TS and the initial acid content C_{A_0}, which are measured in reality with a flow meter and a NIR instrument. Other unmeasured disturbances are simulated by injecting noise into the charge balance equation through C_{Z_0}, i.e. the unknown buffers concentration. The enzymatic reactor has other inputs such as the enzymes stream F_E set by the optimization layer, a recycle input for liquefied fibers F_{FM} (unused), and water F_W (unused).

The structure and the complexity of the L1 adaptive output feedback controller is identical to the one from the pretreatment temperature case. The process is SISO with the pH level as controlled variable (CV) and base flow rate as manipulated variable (MV). There are 3 parameters for tuning: the closed loop desired dynamics m, the filter parameter c, and the adaptation gain Γ, just like in the pretreatment temperature case. The tuning method however is different, and relies on linearizing the model around a nominal operational point and analyzing the closed loop transfer function. The linearization point is chosen around $pH = 5$, which is optimum for cellulosic fibers hydrolysis [11]. The transfer function of the closed loop system dynamics from reference to output are then determined, and the damping ratio ζ is plotted against controller parameters m and c as in the top plot of Figure 5.7. The derivation of the transfer functions are detailed in Paper E. The m and c graph shows that there is a trade-off between choosing these parameters. A fast system with m tending to 1 would require a more aggressive control signal meaning that c needs to be enlarged too for allowing a larger bandwidth. The goal is to stay at the border

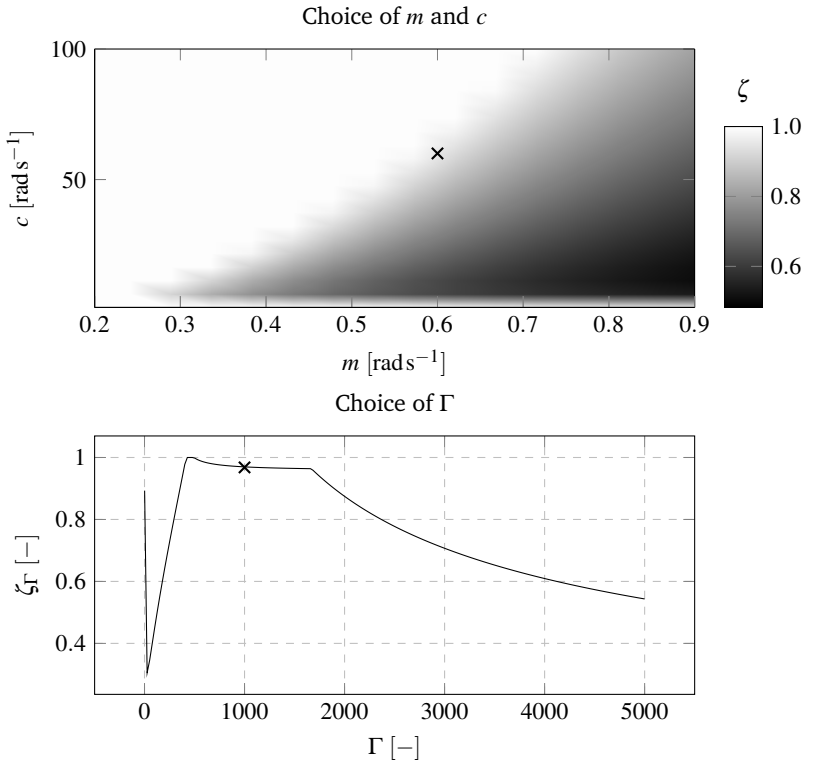
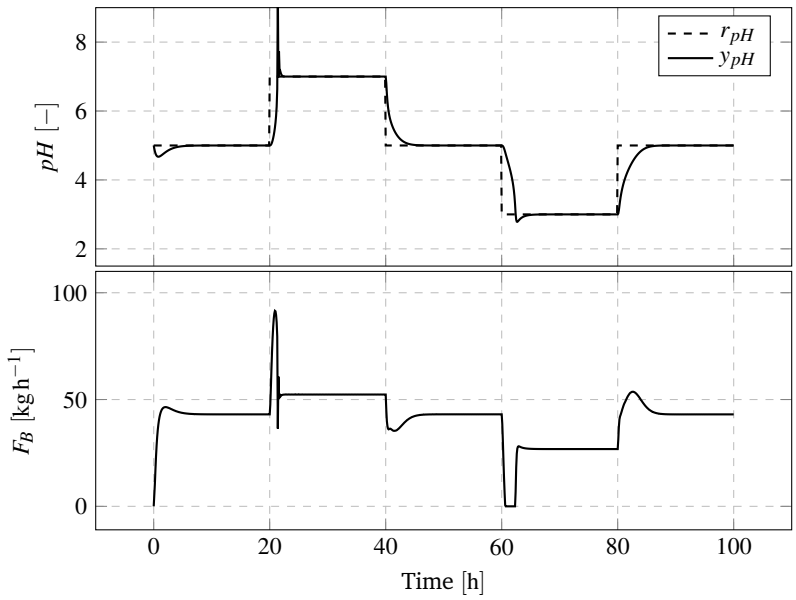


Figure 5.7: Tuning of the L1 adaptive controller.

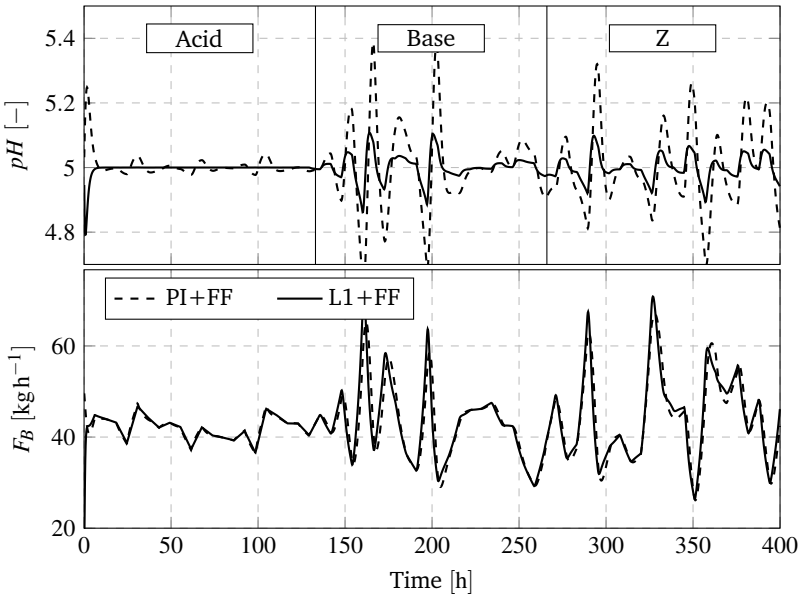
between the white and gray area from the m and c plot.

After choosing m and c, Γ can be set. The closed loop dynamics from reference to $\hat{\sigma}$ are then determined, and the damping ratio is plotted as a function of Γ as in the bottom plot of Figure 5.7. If the adaptation gain is very large then $\hat{\sigma}$ oscillates and introduces noise in the control signal. Γ is chosen such that the damping ratio ζ_Γ of the adaptation law stays relatively close to 1.

The controller is tested for reference tracking of a staircase signal and disturbance rejection with respect to acid, base and unknown buffers. Reference modifications can occur either when type of enzymes change or when a pH profile is tracked, which can happen in fermentation. The reference tracking results are plotted in Figure 5.8(a). In enzymatic hydrolysis the pH level needs to stay in the acidic area, e.g. $pH = 5$, where the mixture is well buffered. At these pH levels, the closed loop system has a reasonable response with little or no overshoots. However, the most challenging area is the neutral zone around $pH = 7$ because the titration curve changes slope rapidly becoming very steep. The neutral zone scenario is more theoretical and is not typical for a real process. The adaptive controller manages to



(a) Reference tracking with an L1 controller. Top plot shows the pH measurement and the reference level while the bottom plot displays the control effort or the addition of base.



(b) Disturbance rejection with an L1 adaptive controller. Top plot shows the reactor pH level in 2 cases: PI+FF and L1 controller with feed-forward (L1+FF). The bottom plot displays the control effort.

Figure 5.8: Reference tracking and disturbance rejection with an L1 adaptive output feedback controller.

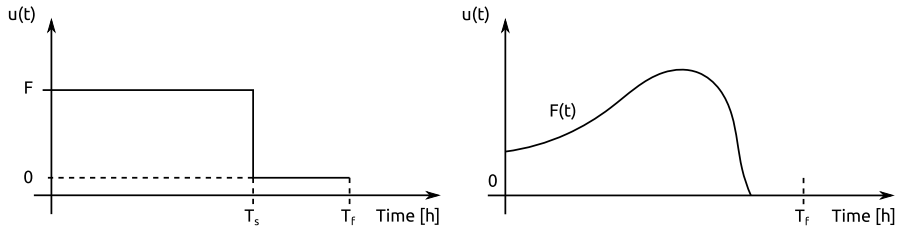
stabilize the process at $pH = 7$ with a large overshoot but converges eventually. As the system is very fast around the neutral zone, a faster adaptation law might be required in order to reduce the overshoot.

The disturbance rejection scenario is shown in Figure 5.8(b), and comprises of three zones: acid, base and unknown buffers Z. Each source of disturbance is enabled only in the designated area. The L1 adaptive controller is compared against a classical Proportional Integral (PI) controller that was tuned using the Skogestad internal model approach for PID tuning [90]. The feed-forward term is enabled for both controllers. The adaptive scheme is not necessarily more aggressive as its control effort is similar to the classical controller but reacts sooner than the PI. Disturbances are successfully rejected in all cases ensuring little deviations from the optimal pH level within a band of ± 0.1 units.

5.4 Optimal Feed Rate Profile for Glucose Fermentation

Figure 4.3(b) from the optimization section showed that fermentation costs can be high in biorefinery operation. The objective of this part of the study is to determine an optimal feed rate profile for liquefied fibers such that the amount of yeast is minimized while achieving glucose depletion in approximately the same amount of time as in a classical operation with a constant feed rate. The results from this section have been developed in collaboration with the Automatic Control Laboratory from École Polytechnique Fédérale de Lausanne (EPFL). A commercial scale fermentation tank can reach a volume of $250\,\mathrm{m}^3$ requiring the fed-batch phase to take 40 h to 60 h to reach the desired hold-up, time when reactions already take place. Plants following a traditional operation fill the fermentation reactors with a constant feed rate. Such a strategy can lead to an early accumulation of inhibitors and sugars that delay the reactions. In order to compensate the inhibition, more yeast would be required, which attracts higher costs.

The two strategies are conceptually illustrated in Figure 5.9. The constant feed



(a) Constant feed rate. (b) Variable feed rate.

Figure 5.9: Constant and variable feed rates.

rate or traditional operation is shown in Figure 5.9(a). The feed rate remains constant at a value F. Once the tank is filled F becomes 0, coinciding to a switching time T_s. Glucose depletion occurs at time T_f. The variable feed rate profile can resemble the shape from Figure 5.9(b), which is characterized by a lower rate in the beginning to avoid inhibitors and sugars accumulation and exponentially increasing as cell biomass grows. Once the tank reaches the set level, the feed rate smoothly becomes 0.

Both scenarios are formulated within the OCP framework [91]. The multi-objective dynamic optimization problems are then solved with ACADO toolkit, which has been successfully utilized previously in other biochemical applications [92, 93]. The decision variables for the constant feed rate optimization problem are set as T_f, T_s, F, and x_{Y_0}, i.e. the final fermentation time, the switching time, the value of the feed rate, and the initial concentration of yeast. The multi-objective cost function aims at minimizing both the yeast seed x_{Y_0} and keep small deviations around the fermentation time T_f. The cost function includes two penalty terms, P_Y and P_T, i.e. the price of yeast and price of fermentation time. The optimization problem is shown next together with its constraints:

$$
\begin{aligned}
\min_{T_f,T_s,F,x_{Y_0}} \quad & x_{Y_0} m(0) P_Y + T_f P_T \\
\text{subject to} \quad & \dot{m}(t) = \begin{cases} F, & 0 \le t < T_s \\ 0, & t \in [T_s, T_f] \end{cases} \\
& \dot{x}(t) = f(x(t), m(t)) \\
& x(0) = x_0 \\
& m(0) = 22.105\,\mathrm{t} \\
& x_Y(0) = x_{Y_0} \\
& m(T_f) = 220\,\mathrm{t} \\
& x_G(T_f) \le 0.1\,\mathrm{g\,kg^{-1}} \\
0.0 \le\ & x_F(t) \le 0.05\,\mathrm{g\,kg^{-1}} \\
0.0 \le\ & F \le 10\,\mathrm{t\,h^{-1}} \\
30 \le\ & T_f \le 90\,\mathrm{h} \\
0 \le\ & T_f - T_s \\
0.0 \le\ & x_{Y_0} \le 12\,\mathrm{g\,kg^{-1}}
\end{aligned}
\tag{5.2}
$$

The initial mass of yeast is calculated as x_{Y_0} times the initial reactor mass, i.e. $m(0)$. Yeast is grown locally with costs reflected in term P_Y.

The first dynamic constraint refers to mass accumulation in time:

$$
\dot{m}(t) = \begin{cases} F, & 0 \le t < T_s \\ 0, & t \in [T_s, T_f] \end{cases} \tag{5.3}
$$

The feed rate switches to 0 at time T_s.

The second dynamic constraint refers to the fermentation reactor kinetics, which were presented in Chapter 3:

$$\dot{x}(t) = f(x(t), m(t), F) \tag{5.4}$$

where x symbolizes the species concentrations, and m is the reactor hold-up.

The initial conditions constraints consists of: the fiber concentrations when the fed-batch phase starts, i.e. $x(0) = x_0$, initial reactor mass $m(0) = 22.105\,\mathrm{t}$, and yeast seed $x_Y(0) = x_{Y_0}$. The final time constraints refer to reactor fill limit $m(T_f) = 220\,\mathrm{t}$, and glucose depletion $x_G(T_f) \leq 0.1\,\mathrm{g\,kg^{-1}}$. Accumulation of inhibitors at any given time is limited by $x_F(t) \leq 0.05\,\mathrm{g\,kg^{-1}}$. Additional bounds are set for all decision variables.

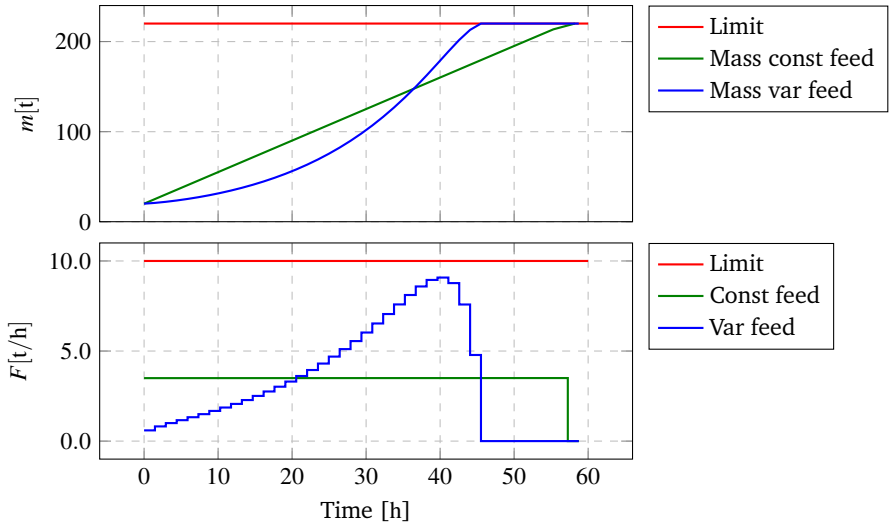
The optimization control problem for variable feed profile is formulated below:

$$\begin{aligned}
\min_{F(t),T_f,x_{Y_0}} \quad & x_{Y_0} m_0 P_Y + T_f P_T \\
\text{subject to} \quad & \dot{x}(t) = f(x(t), F(t)) \\
& x(0) = x_0 \\
& m(0) = 22.105\,\mathrm{t} \\
& m(T_f) = 220\,\mathrm{t} \\
& x_Y(0) = x_{Y_0} \\
& x_G(T_f) \leq 0.1\,\mathrm{g\,kg^{-1}} \\
0.0 \leq\ & x_F(t) \leq 0.05\,\mathrm{g\,kg^{-1}} \\
0.0 \leq\ & F(t) \leq 10\,\mathrm{t\,h^{-1}} \\
30 \leq\ & T_f \leq 90\,\mathrm{h} \\
0.0 \leq\ & x_{Y_0} \leq 12\,\mathrm{g\,kg^{-1}}
\end{aligned} \tag{5.5}$$

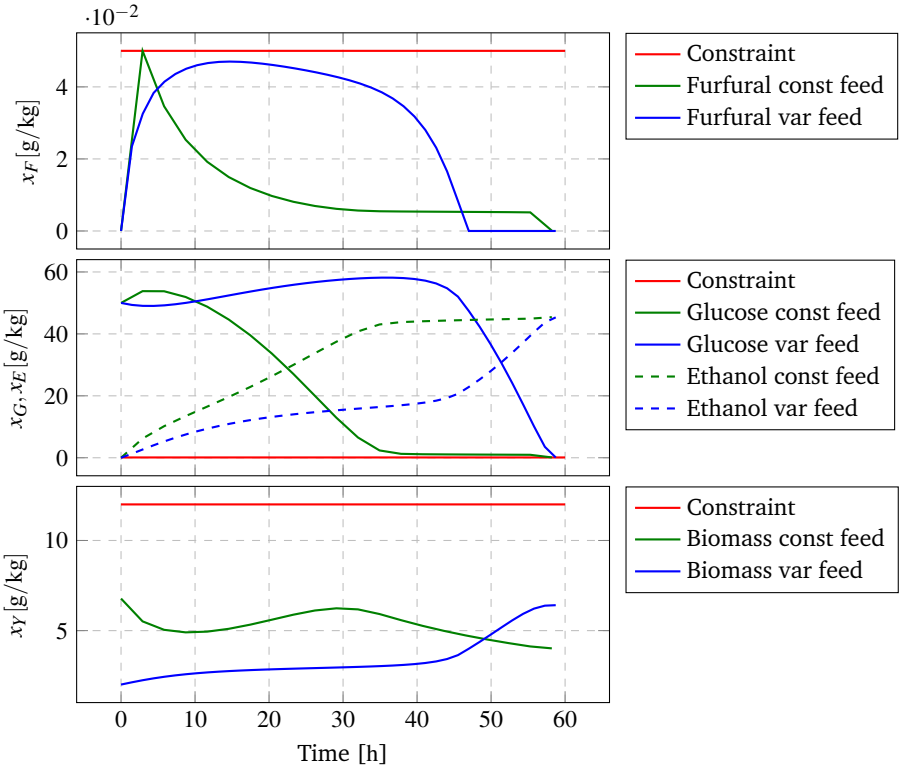
The problem is similar to OCP (5.2) except that the decision variable $F(t)$ becomes a function of time and switching time T_s is removed. The cost function and the constraints are identical.

The results of the two strategies are compared in Figure 5.10. The mass and the feed rates are illustrated in Figure 5.10(a). The constant inflow of liquefied fibers is set to approximately $3.85\,\mathrm{t\,h^{-1}}$ and fills the reactor in about 57 h. The variable feed rate starts at a lower level than in the constant feed case but gradually increases as time passes until it reaches a peak of $9\,\mathrm{t\,h^{-1}}$ around 40 h. In the next 5 h the feed rate decreases fast until it reaches $0\,\mathrm{t\,h^{-1}}$. The tank is filled faster than in the constant rate case by about 12 h.

The reactor kinetics are shown in Figure 5.10(b). The constant feed rate is limited by the inhibitors accumulation constraints, which is reached in the first 3 h as shown in the top plot. In the feed profile case, the flow rate allows accumulation of inhibitors near the constraint limit. The system does not operate exactly at the



(a) Mass and feed rate.



(b) Furfural, glucose, ethanol and cell biomass.

Figure 5.10: Comparison between constant and variable feed rates.

constraint limit because a high feed rate also accumulates glucose, which inhibits ethanol production. Glucose and ethanol concentrations can be observed in the middle plot. Glucose remains in the reactor for a longer time in the variable feed rate case and its depletion is accelerated towards the end of the fermentation process. The ethanol yield is identical in both scenarios and is reached in the same amount of time. However, cell biomass growth is substantially improved because the variable feed rate requires a much lower initial yeast seed.

The OCP solutions and the costs of the two strategies are compared in Table 5.1. The variable feed profile strategy shows a great potential for saving on yeast seed, which is reduced with more than 50 % in this scenario.

Table 5.1: Comparison of costs between the two feed strategies.

	Constant feed	Variable feed
Objective function	14962	10241
T_f	58.20	58.73
x_{Y_0}	6.77	2.0082
T_s	57.14	-
F	3.5	-
Mass of yeast (kg)	136	40
Potential savings: 96 kg of yeast/batch (4800 dkk/batch)		

Determining the input profiles requires a mathematical model of the process in order to predict its behavior in time. However, in reality models have uncertainties and a Real-Time Optimization (RTO) scheme needs to be constructed. RTO assumes on-line model adaptation, which might prove to be difficult to implement in reality for this case. A better approach is to exploit the Necessary Conditions for Optimality (NCO) and select appropriate control variables for tracking NCOs rather than performing numerical optimization of a complex problem [94, 95]. From the above results, it can be seen that the system operates near the inhibitors constraint, i.e. furfural concentration limit. If the process would have a measurement of inhibitors then a feed rate controller can be added for determining a better feed profile than a constant one, which would save costs on yeast.

5.5 Conclusions

This chapter presented two applications of the L1 adaptive output feedback controller: one for biomass pretreatment temperature and another one for enzymatic pH. The temperature controller setpoint is connected to the optimization layer

developed in the previous chapter, while the pH controller ensures the efficiency of the enzymatic activity. The main presented novelties came from the tuning methods of these controllers: one based on closed loop transfer function analysis, and another one based on formulating an optimization problem.

The adaptive controllers were followed by an optimal controller for determining the feed rate profile in glucose fermentation. These results showed that there is a high potential in yeast economy that can be achieved with a variable feed rate profile. Further investigation is required in order to transform the simulation results into an implementable RTO scheme.

Chapter 6

Conclusions and Future Research

6.1 Summary of Conclusions

The core contribution of this research is a complex dynamic model library suitable for industrial size plants. Other parts of this work include model-based tools that improve the operation of large scale biorefineries: an optimization layer for maximizing the economical profit of the biorefinery, and an advanced control layer to support an optimal operation.

The dynamic model library captured in great detail and accurately the behavior of the key steps in biomass conversion, i.e. hydrothermal pretreatment with steam and enzymatic hydrolysis. The library was completed with co-fermentation supported by GMO yeast. The models from this work were subject to a comprehensive analysis methodology for validation, and for assessing their sensitivity and uncertainty. This work published for the first time validation against demonstration scale real data, which proves that the model library is ready to be used in real industrial applications.

The models were then exploited to optimize the biorefinery operation for maximizing its economical profit. The plant was treated in an integrated manner in order to capture the trade-offs between the biomass conversion steps. Too little pretreatment might not open the biomass sufficiently leading to an increase in enzyme dosage and yeast seed while an over pretreatment lead to inhibitors formation that decrease the ethanol yield in fermentation. The optimization layer identifies the best compromise, and offers setpoints for the key process parameters, i.e. thermal reactor temperature, enzyme dosage and yeast seed. The optimization part of the study was carried in a simulated environment and showed potential economical improvements of 18 % over a traditional plant operation.

The optimization layer is supported by an advanced automation and control layer. Pretreatment temperature and pH in enzymatic hydrolysis are key parameters

of the biorefinering process. Two high performance L1 adaptive controllers were developed: one for pretreatment temperature and another one for pH in enzymatic hydrolysis. The L1 adaptive controller for the thermal reactor temperature ensured little deviations from the setpoint when the particle pump extracted steam for pressurization of biomass. Changes in setpoints as indicated by the optimization layer can also be performed with a reasonable fast first order response. The enzymatic pH posed a higher challenge due to its nonlinear titration curve. Once again the L1 adaptive controller was capable of minimizing deviations around the setpoint ensuring an efficient enzymatic activity. The L1 adaptive controllers were tuned systematically: by formulating an optimization problem that minimizes the IAE, and by analyzing closed loop transfer functions for obtaining a specific performance. The last component added to the control layer was an optimum feed rate profile of liquefied fibers to fermentation. The challenge was to avoid accumulation of inhibitors and sugars that could delay reactions and decrease ethanol yield. The results showed an important reduction in yeast amount for obtaining the same ethanol yield.

6.2 Future Research

There is great potential for future research in the area of biorefinery automation. The advanced control and optimization tools formulated in this work offer a higher profitability to the Inbicon technology making it more attractive to customers. As future perspectives, emphasis will be placed on implementation on real-time hardware making all these tools part of the commercial Inbicon technology. The modeling work is continuous and will be maintained and updated with newly documented phenomena, as well as validated against more real data.

Simultaneous saccharification and fermentation can be further improved. The method used for determining the optimal feed rate profile can be extended for pH and temperature because enzymes and yeast do not have the same optimal conditions. This would lead to an optimal economic operation of the fermentation process that would save even more costs.

The optimization work showed great potential and should be extended with real-time optimization for implementation possibly following NCO tracking methods. The dynamic models can also be transformed into soft sensors and state observers such that the refinery operators can monitor all variables of interest even in the absence of a real measurement. Faults happen in reality and detection and isolation are important for an efficient operation. A model-based diagnosis layer would add even further value to the Inbicon technology.

Paper A

Dynamic Modeling and Validation of a Biomass Hydrothermal Pretreatment Process - A Demonstration Scale Study

Remus Mihail Prunescu[1], Mogens Blanke[1], Jon Geest Jakobsen[2], Gürkan Sin[*3]

[1]Department of Electrical Engineering, Automation and Control Group, Technical University of Denmark, Elektrovej Building 326, 2800, Kgs. Lyngby, Denmark

[2]Department of Process Control and Optimization, DONG Energy Thermal Power A/S, Nesa Allé 1, 2820, Gentofte, Denmark

[3]CAPEC-PROCESS, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads Buildings 227 and 229, 2800, Kgs. Lyngby, Denmark

Abstract:
Hydrothermal pretreatment of lignocellulosic biomass is a cost effective technology for second generation biorefineries. The process occurs in large horizontal and pressurized thermal reactors where the biomatrix is opened under the action of steam pressure and temperature to expose cellulose for the enzymatic hydrolysis process. Several by-products are also formed, which disturb and act as inhibitors downstream. The objective of this study is to formulate and validate a large scale hydrothermal pretreatment dynamic model based on mass and energy balances, together with a complex conversion mechanism and kinetics. The study includes a comprehensive sensitivity and uncertainty analysis, with parameter estimation from real-data in the 178 °C to 185 °C range. To highlight the application utility of the model, a state estimator for biomass composition is developed. The predictions capture well the dynamic trends of the process, outlining the value of the model for simulation, control design, and optimization for full-scale applications.

[*]Principal corresponding author. Tel.: +45 45252806; E-mail: gsi@kt.dtu.dk

A.1 Introduction

Lignocellulosic biomass, e.g. wheat straw, corn stover, bagasse etc, consist of cellulose, hemicellulose, lignin, ash, and a negligible amount of residues [37]. The cellulosic fibers contain glucose units, which are necessary for biofuel production, but layers of hemicellulose and lignin make cellulose hardly accessible. The goal of the pretreatment process is to relocate lignin, and partially hydrolyze the hemicellulose, which opens the biomatrix for cellulose such that enzymes can easily access it in the enzymatic hydrolysis process downstream [40].

Chiaramonti et al. [44] review various methods of pretreatment, e.g. autohydrolysis, steam explosion, acid hydrolysis, alkaline hydrolysis, and many others. Studies show that hydrothermal pretreatment with steam excels in cost effectiveness and, therefore, has been commercialized in large scale second generation biorefineries. Integrating the biorefinery with a power plant following the Integration Biomass Utilization System (IBUS) also contributes to reducing costs [4]. The pretreatment process partially depolymerizes hemicellulose creating several degradation and by-products, i.e. organic acids, xylooligomers, xylose, and inhibitors, e.g. furfural and 5-HMF, which impact the downstream processes. The organic acids, i.e. acetic, succinic and lactic acid influence the pH of pretreated fibers and become an issue in the enzymatic hydrolysis process [6]. Xylooligomers and xylose act as strong inhibitors of cellulose hydrolysis by enzymes [51], while furfural and 5-HMF inhibit the fermentation process [52]. Also carbohydrates react with degradation products such as furfural to create spherical droplets with lignin like structure named pseudo-lignin, which can degrade the enzymatic activity [53]. Experimental studies show that reactor temperature and retention time relate to biomass conversion [50].

Lavarack et al. [73] formulate a mechanistic acid hydrolysis model capable of predicting cellulose, xylan and furfural concentrations but no one evaluated the model for steam pretreatment, and at a large scale. This model has been used in later studies for model-based optimization of bioprocesses under uncertainty [9] and biorefinery configurations [96]. Overend et al. [97] present an empirical modeling alternative known as the severity factor, which Petersen et al. [50] validated in laboratory experiments for xylan recovery and furfural formation. These models are incomplete because: (1) they miss production of important by-products such as organic acids, xylooligomers, and pseudo-lignin; and (2) assume a uniform thermal environment, which is not the case in a full scale reactor [71]. This study extends the existing Lavarack et al. [73] model to demonstration scale using computational fluid dynamics techniques, taking into account temperature variations in a large scale thermal reactor, and production of the above enumerated by-products. The

model is then calibrated and validated against real data that were logged throughout several hours of operation at a demonstration scale facility.

This study also assesses the model reliability through a comprehensive sensitivity and uncertainty analysis. The sensitivity analysis quantifies the importance of each model parameter and creates an identifiable subset of parameters that is used for parameter estimation following the methodology from Sin et al. [8] and Prunescu and Sin [6].

Samples of fibers were collected after the pretreatment process and were analyzed with near infra-red instruments (NIR) to determine their composition. The model parameters are estimated using the NIR readings. The uncertainty analysis determines a confidence interval for model predictions and is carried with respect to both model and feed parameters following the method from Sin et al. [8]. As a global sensitivity measure, the standardized regression coefficients (SRC) are computed in order to identify the model parameters responsible for most of the variations in model predictions [66]. A residual analysis follows to identify how much of the signal is represented by the model. The study ends with a model application as a state estimator by using a static Kalman filter. The state estimator not only filters the NIR measurements but also predicts by-products formation such as C5 sugars and inhibitors.

This study has the following structure: the *Methods* section gives an overview of a state of the art demonstration scale biorefinery from where the real measurements were collected; a *Model Development* section follows, which formulates the mathematical model for the pretreatment process along with the model analysis methodology; the *Results and Discussion* section presents the model analysis and validation results, and the model application as a state estimator; the study ends with a *Conclusions* section, which summarizes all findings.

A.2 Methods

A.2.1 Biorefinery Experimental Setup

Figure A.1 shows an overview of a demonstration scale second generation bioethanol plant. Biomass is first pretreated in a continuous thermal reactor where the temperature can be set within 160 °C to 210 °C, and the retention time can be adjusted from 6 min to 20 min. Following the IBUS concept, a nearby power plant supplies the biorefinery with steam for pressurizing the thermal reactor, thus reaching the necessary reaction temperatures. The steam supply pipe provides saturated steam at 18×10^5 Pa. After the thermal reactor, the pretreated slurry is first washed and then the liquid part is separated from fibers in a screw press. The liquid part is

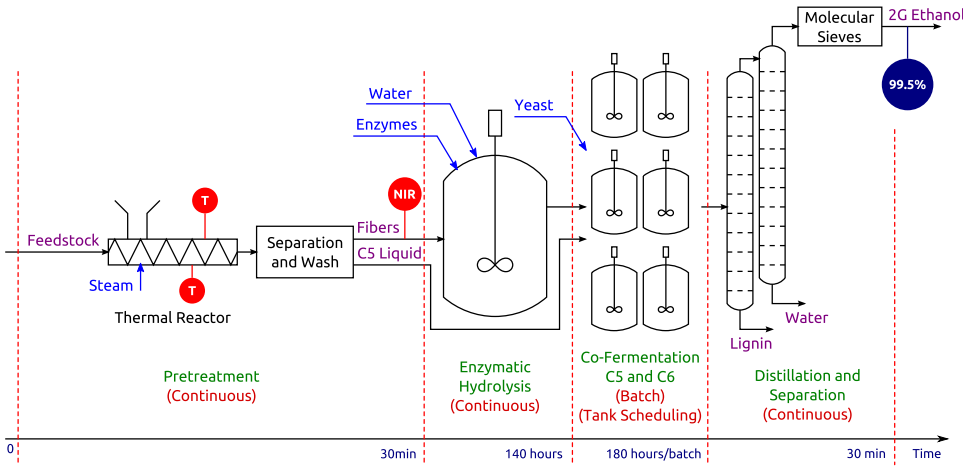


Figure A.1: Process flow diagram for a second generation biorefinery: pretreatment, enzymatic hydrolysis, C5 and C6 co-fermentation, and purification. The samples for NIR analysis are extracted after the pretreatment process every 10 min. The thermal reactor is monitored with two temperature sensors: one at the top and another one at the bottom of the reactor.

rich in C5 sugars, i.e. xylooligomers, xylose and arabinose, as a result of partially hydrolyzing the hemicellulose in the thermal reactor.

The cellulosic fibers are transported on a conveyor belt to the enzymatic hydrolysis tank. Enzymes are added in the liquefaction tank where cellulose and the remaining xylan are converted to glucose and xylose. The enzymatic hydrolysis process has been detailed and analyzed in Prunescu and Sin [6]. The C5 and C6 sugars are then co-fermented for ethanol production in scheduled batch reactors with genetically modified organisms (GMOs) for enhancing bioethanol production.

The purification and separation phase contains two distillation columns and molecular sieves. Lignin is separated in the first distillation column, while ethanol is purified to 99.5 % in the second column and in the molecular sieves. The recovered lignin is transported to an evaporation plant and solidified as bio-pellets, which are sent to the nearby power plant for burning.

There is a timeline indicator at the bottom of Figure A.1 showing the retention time for each section of the biorefinery. The pretreatment process and distillation are the fastest processes with a duration of maximum half an hour, while the enzymatic hydrolysis and fermentation can last 5 to 7 days each.

The demonstration scale facility has a processing capacity of $4000\,\mathrm{kg\,h^{-1}}$ of biomass [3]. Samples of pretreated fibers were extracted after the pretreatment process every 10 min for a total duration of 15 h. The samples were then analyzed

with near infra-red instruments (NIR) to determine their composition with respect to cellulose, xylan, lignin, acetic acid, and furfural. The thermal reactor was monitored with two temperature sensors, one placed at the top of the tank for measuring the temperature in the steam layer, and another one placed at the bottom of the reactor to measure the biomass temperature.

A.2.2 Dataset

Figure A.2 shows the raw temperature and NIR measurements collected throughout the 15 h of operation. The dataset is split in two subsets: one is used for estimation purposes, while the other for validation. The estimation dataset is identified in all figures with a grey background. The top reactor temperature oscillates within 180 °C to 190 °C due to the pressurization unit, which takes the biomass from atmospheric to reactor pressure with recycled steam from the reactor. The whole process is thoroughly explained in Prunescu et al. [71]. There is a temperature difference between the top and bottom of the reactor due to the insulation properties of the biomass, and also because vertical mixing in the reactor was not possible due to a low horizontal pushing speed.

A.2.3 Model Analysis Methodology

The model analysis follows the methodology from Table A.1:

1. The first step is to calibrate the model considering the entire set of parameters. Ideally, this system identification exercise should give the set of parameters that has the smallest sum of squared errors between model predicted output and actual measurements. In the present case, this is a nonlinear least squares problem and local minima can be obstacles. The parameter estimation may be solved in `Matlab` with the `fminsearch` function, coupled with a cost function that calculates the prediction error, if the initial parameter guess is sufficiently close to a global minimum.

2. The second step is to investigate which parameters of the model could be determined given the input and the model structure [63]. This selection is achieved through assessment of sensitivity of the partial derivatives of the cost function with respect to each model parameter. After computing the sensitivities, δ^{msqr}, all parameters are ranked with respect to their value of δ^{msqr}. Parameters that have low sensitivity are more uncertain that those with high sensitivity and would not contribute to model accuracy. Therefore, a

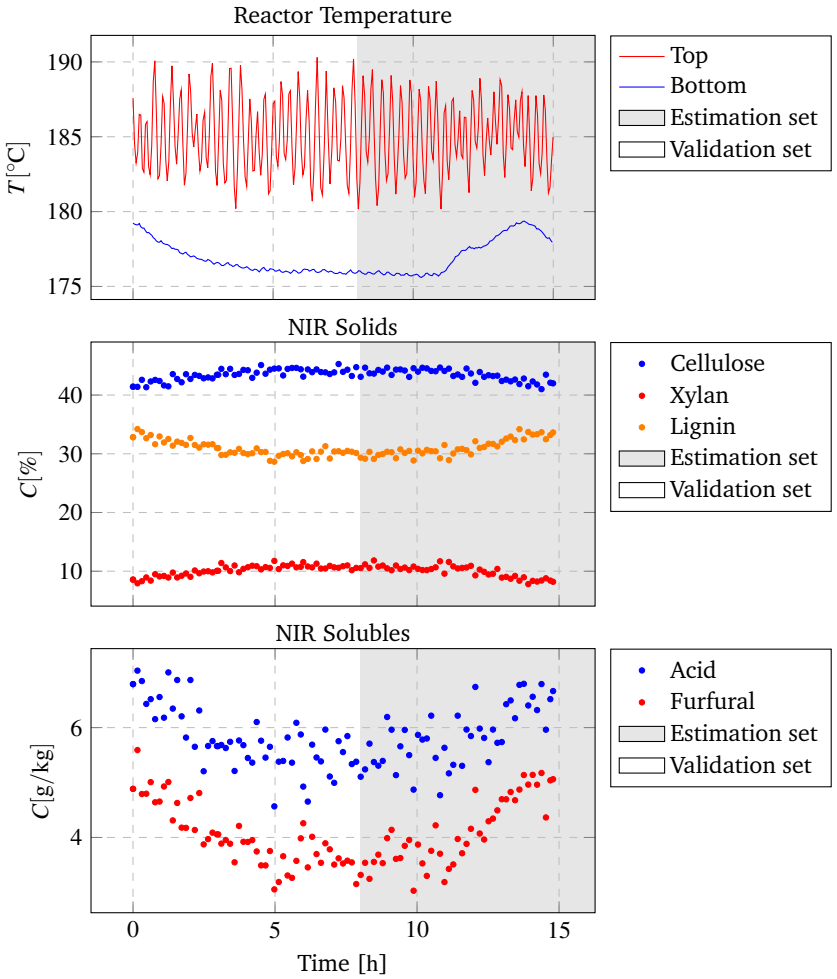


Figure A.2: The raw dataset. The top plot shows the reactor temperatures measured by the top and bottom sensors. NIR offers information on the solid and soluble content of pretreated fibers. The whole dataset is split into estimation and validation subsets.

relevant subset of parameters is selected based on δ^{msqr} being higher than a threshold.

3. In the third step the reduced set of parameters is identified using the NIR measurements from the demonstration scale plant. The correlation matrix and standard deviations of the estimates are also computed.

4. This step quantifies the prediction uncertainty. Having the covariance matrix and standard deviations from the previous step allows Latin Hypercube Sampling (LHS) [65] with correlation control. The feed parameters is another source of uncertainty and is included in this analysis. N Monte Carlo simulations are then run with sampled values and the 5th-95th percentiles of the model predictions are found. A global sensitivity analysis follows by fitting a linear model from parameters to model predictions and the standardized regression coefficients are computed to identify which parameters are the most important for explaining the output uncertainty.

5. The model estimation error or the residuals are analyzed in this step. A simulation is run with the estimated parameters using the entire set of data (not only the estimation set). The residuals distribution and autocorrelation are calculated in order to assess the quality of model predictions. A good model captures most of the signal in measurements and is characterized by residuals being Gaussian with uncorrelated increments.

A.3 Model Development

The mathematical model consists of mass and energy balances for the pretreatment process. In large scale plants, the most common continuous thermal reactor is a long tank with cylindrical shape. This study employs simplified computational fluid dynamics tools for modeling the composition and temperature profiles.

A.3.1 Mass Balance

The thermal reactor has a continuous operation and the mass balance is established as the accumulation of mass per unit of time equals the difference between inflow and outflow rates:

$$\frac{dM}{dt} = F_{in} - F_{out} \tag{A.1}$$

where M is the total mass of biomass inside the reactor, F_{in} is the inflow rate of pressurized biomass and F_{out} is the outflow rate of pretreated biomass.

Table A.1: Sensitivity and uncertainty analysis methodology. The output from step $k-1$ is the input to step k.

#	Step	Description	Output
1	Model initialization	• Initialization of all model parameters to obtain a good working model fit;	θ_0
2	Sensitivity analysis	• List of significant parameters; • Find an identifiable parameter subset.	δ^{msqr} θ_{R_0}
3	Parameter estimation	• Identify parameter subset; • Correlation matrix; • Confidence interval for parameters.	$\hat{\theta}_R$ $\mathbf{R}_\theta$ σ
4	Uncertainty analysis	• Calculate prediction uncertainty of the model; • Sensitivity analysis with standardized regression coefficients.	5th-95th percentile β
5	Residual analysis	• Run simulation with the estimated parameters and using the entire dataset • Check probability distribution of model estimation errors or residuals • Compute the autocorrelation function	

A.3.2 Composition Balance

Pretreated fibers contain the following species: cellulose, xylan, arabinan, lignin, acetyl groups, ash, glucose, xylooligomers, xylose, organic acids, furfural, 5-HMF, and other components in negligible amounts. The change of species concentration with respect to time is a combination of convection and diffusion effects plus production and consumption terms, which is modeled with the generic Convection-Diffusion-Reaction equation [70]:

$$\frac{\partial C}{\partial t} = -\nabla (vC) + \nabla (D \nabla C) + R \tag{A.2}$$

C is the species concentration vector, v is the horizontal transportation speed, ∇ is the gradient operator, D is the diffusion coefficient, and R is the reaction rate vector. Most of the particles are in solid state and they do not diffuse. Solubles, i.e. xylooligomers, xylose, arabinose, organic acids, furfural and 5-HMF, diffuse but the

snail pushing the biomass along the reactor separates the tank in chambers due to its pallets, which leads to little exchange of matter between these chambers. If the reactor is split in the same number as the chambers created by the rotational snail pallets then the diffusion coefficient can be dropped:

$$\frac{\partial C}{\partial t} = -\nabla (vC) + R \tag{A.3}$$

The horizontal speed v is set to a constant value such that to meet a certain retention time:

$$v = \frac{L_r}{t_r} \tag{A.4}$$

where L_r is the reactor length and t_r is the retention time. E.g., a thermal reactor of 12 m with a retention time set to 15 min has a horizontal speed of $0.013\,\mathrm{m\,s^{-1}}$, which is a low speed not capable of agitating the biomass vertically.

Since the reactor tank is longer on one axis, the partial derivatives with respect to the other axis could be dropped. It is assumed that the biomass composition changes only along the reactor and remains uniform in the other directions:

$$\frac{\partial C}{\partial t} = -v\frac{dC}{dz} + R \tag{A.5}$$

where z is the axis oriented along the reactor from inlet to outlet. Due to the low horizontal speed, the vertical uniformity assumption stands valid only if the biomass level in the reactor is low. At high biomass levels, the temperature gradient causes different pretreatment layers and the reactor needs to be split on the vertical axis too [71]. The data in this study was collected at low reactor levels and vertical uniformity can be assumed but differences between the steam and biomass layer temperatures still exist.

Figure A.3 illustrates the concept of splitting a reactor tank of length L_r into N smaller reactors, or cells, of width δ_z. Equation A.5 is then discretized in space using a Upwind Difference Scheme (UDS), which is a more stable technique for moving environments [72]:

$$\frac{dC_k}{dt} = \frac{u_z}{\delta_z}(C_{k-1} - C_k) + R_k \tag{A.6}$$

C_k is the composition vector in central cell k, C_{k-1} is the composition vector from the western neighbor, and R_k is the reaction rate vector from current cell k. Movement from west to east (left to right) is assumed. The composition vector will be detailed in the next section.

A.3.3 Mechanistic Modeling for Hydrothermal Mediated Pretreatment

Raw biomass consists of cellulose, hemicellulose (arabinan and xylan), lignin, acetyl groups, ash, water, and other components in negligible amounts. Table A.2

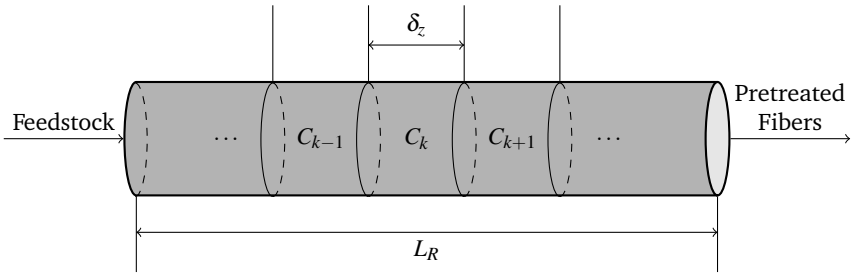


Figure A.3: Computational fluid dynamics methodology for the thermal reactor. The reactor tank is split into N cells of equal width δ_z. A cell C_k has two neighbors, i.e. C_{k-1} to the west, and C_{k+1} to the east. Biomass is transported from left to right. The length of the reactor is L_r.

shows the composition fractions of different types of lignocellulosic biomass.

Table A.2: Composition of different raw biomass types in percentage of dry matter. Some components were not measured (n/m).

Biomass	Cellulose	Xylan	Arabinan	Lignin	Ash	Acetyls	Other	Reference
Bagasse	39.0	21.8	1.8	24.8	3.9	3.3	5.4	[43]
Wheat straw	39.8	24.5	2.8	22.6	4.2	n/m	6.1	[40]
Wheat straw	42.1	23.4	1.9	21.6	2.1	n/m	8.9	[41]
Miscanthus	38.2	20.9	1.5	26.4	2.6	4.1	6.3	[44]
Corn stover	35.0	18.5	n/m	13.9	3.8	3.2	25.6	[42]
Corn stover	34.0	19.2	2.5	12.3	4.7	2.9	24.4	[43]

Figure A.4 illustrates the reaction paths that occur in the thermal reactor. There is little change to cellulose, which is recovered almost entirely in the fiber fraction after the pretreatment process under optimal conditions [50]. However, a small fraction of cellulose could be hydrolyzed to glucose, which is further degraded to 5-HMF. Arabinan is completely hydrolyzed to arabinose [40]. The xylan to xylose path has the xylooligomers intermediate product, which is important to predict because it acts as a very strong inhibitor in the enzymatic hydrolysis process along with xylose [51]. Pentoses, i.e. arabinose and xylose, are further decomposed into furfural and other degradation products. Carbohydrates, i.e. xylooligomers, xylose, arabinose, and glucose, react with furfural and 5-HMF to create spherical droplets with a lignin like structure, also known as pseudo-lignin [53]. Furfural and 5-HMF production is important to monitor due to its inhibitory effects on fermentation [52], while organic acids influence the pH of the enzymatic hydrolysis process [11]. Pseudo-lignin can degrade the enzymatic activity [53].

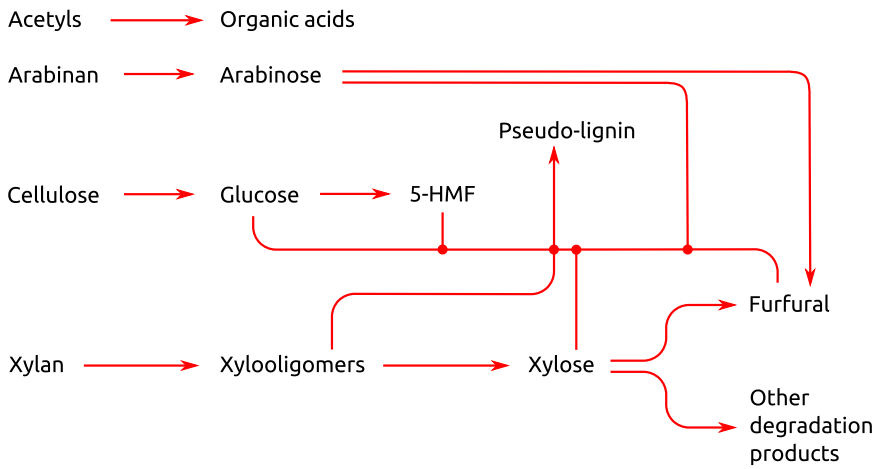


Figure A.4: Hydrothermal pretreatment conversion mechanism.

Reaction rates are modeled as first order Arrhenius temperature dependent equations. Glucose production rate is defined as:

$$r_G = k_G \exp\left(-\frac{E_G}{R_g \cdot T_K}\right) C_{C_S} \tag{A.7}$$

where r_G is the glucose production rate, k_G is the reaction constant, E_G is the activation energy, R_g is the universal gas constant, T_K is the environment temperature in Kelvin degrees, and C_{C_S} is the concentration of cellulose (solid), or the substrate.

Glucose degradation rate to 5-HMF or r_H is modeled similarly with concentration of glucose C_G as substrate:

$$r_H = k_H \exp\left(-\frac{E_H}{R_g \cdot T_K}\right) C_G \tag{A.8}$$

Arabinan hydrolysis forms arabinose following the below rate expression:

$$r_A = k_A \exp\left(-\frac{E_A}{R_g \cdot T_K}\right) C_{A_S} \tag{A.9}$$

where C_{A_S} is the concentration of arabinan (solid).

Xylan degradation produces xylooligomers with rate r_{Xo}, which further decompose to xylose with rate r_X defined as follows:

$$r_{Xo} = k_{Xo} \exp\left(-\frac{E_{Xo}}{R_g \cdot T_K}\right) C_{X_S} \tag{A.10}$$

$$r_X = k_X \exp\left(-\frac{E_X}{R_g \cdot T_K}\right) C_{Xo} \tag{A.11}$$

where C_{X_S}, and C_{Xo} are concentrations of xylan (solid), and xylooligomers (liquid).

Furfural is produced from pentoses, i.e. from both arabinose and xylose as follows:

$$r_F = k_F \exp\left(-\frac{E_F}{R_g \cdot T_K}\right)(C_X + C_A) \tag{A.12}$$

where C_X and C_A are concentrations of xylose (liquid) and arabinose (liquid). The amount of furfural produced from xylose is denoted as r_{F_X}, and the one from arabinose as r_{F_A}.

Furfural and 5-HMF bond to carbohydrates to create pseudo-lignin [53], which is modeled in this study in the following way:

$$r_L = k_L \exp\left(-\frac{E_L}{R_g \cdot T_K}\right)(C_{Xo} + C_X + C_A + C_G)(C_F + C_H) \tag{A.13}$$

where C_F is the concentration of furfural (liquid) and C_H is the concentration of 5-HMF (liquid). Equation (A.13) can be split into:

$$\begin{aligned}
r_{L_{Xo}} &= k_L \exp\left(-\frac{E_L}{R_g \cdot T_K}\right) C_{Xo}(C_F + C_H) \\
r_{L_X} &= k_L \exp\left(-\frac{E_L}{R_g \cdot T_K}\right) C_X(C_F + C_H) \\
r_{L_A} &= k_L \exp\left(-\frac{E_L}{R_g \cdot T_K}\right) C_A(C_F + C_H) \\
r_{L_G} &= k_L \exp\left(-\frac{E_L}{R_g \cdot T_K}\right) C_G(C_F + C_H)
\end{aligned} \tag{A.14}$$

which denote pseudo-lignin produced from xylooligomers, xylose, arabinose and glucose when they bind to both furfural and 5-HMF. Then Equation (A.13) becomes:

$$r_L = r_{L_{Xo}} + r_{L_X} + r_{L_A} + r_{L_G} \tag{A.15}$$

Equation (A.13) can also be split into:

$$\begin{aligned}
r_{L_F} &= k_L \exp\left(-\frac{E_L}{R_g \cdot T_K}\right)(C_{Xo} + C_X + C_A + C_G)C_F \\
r_{L_H} &= k_L \exp\left(-\frac{E_L}{R_g \cdot T_K}\right)(C_{Xo} + C_X + C_A + C_G)C_H
\end{aligned} \tag{A.16}$$

where r_{L_F} is the production rate of pseudo-lignin with furfural participation, while in r_{L_H} 5-HMF participates.

Acetyls are released during hemicellulose hydrolysis and lead to organic acids formation with rate r_{Ac}:

$$r_{Ac} = k_{Ac} \exp\left(-\frac{E_{Ac}}{R_g \cdot T_K}\right) C_{Ac_S} \tag{A.17}$$

where C_{Ac_S} is the concentration of acetyl groups in the hemicellulose (solid).

The composition vector C_k from Equation (A.6) contains all components from the mechanistic scheme. The reaction rates from this section are put into a reaction rates vector R_k. C_k and R_k are shown next in vector form:

$$C_k = \begin{bmatrix} C_{C_S} \\ C_{X_S} \\ C_{A_S} \\ C_{L_S} \\ C_{Ac_S} \\ C_G \\ C_{Xo} \\ C_X \\ C_A \\ C_{Ac} \\ C_F \\ C_H \\ C_W \\ C_O \end{bmatrix} \quad R_k = \begin{bmatrix} -r_G \\ -r_{Xo} \\ -r_A \\ r_L \\ -r_{Ac} \\ r_G - r_{O_G} - (1-\alpha) r_{L_G} \\ r_{Xo} - r_X - (1-\alpha) r_{L_{Xo}} \\ r_X - r_{F_X} - r_{O_X} - (1-\alpha) r_{L_X} \\ r_A - r_{O_A} - r_{F_A} - (1-\alpha) r_{L_A} \\ r_{Ac} \\ r_F - \alpha r_{L_F} \\ r_H - \alpha r_{L_H} \\ 0 \\ r_{O_X} + r_{O_G} + r_{O_A} \end{bmatrix} \tag{A.18}$$

where C_W is the water content, and α is a stoichiometric parameter for furfural and 5-HMF participation in pseudo-lignin formation. In order to close the mass balance, the sum of all elements in R_k has to be 0, and the sum of all elements in vector C_k is $1000\,\mathrm{g\,kg^{-1}}$ at any time t:

$$\sum R_k = 0\,\mathrm{g/(kgs)} \qquad \sum C_k = 1000\,\mathrm{g\,kg^{-1}} \tag{A.19}$$

A.3.4 Energy Balance

The steam layer energy balance together with a temperature controller for the thermal reactor have been formulated in Prunescu et al. [12]. The energy balance for the biomass layer has been studied in Prunescu et al. [71] and is simplified in this paper by a distributed parameters model on one axis, which is discretized along the reactor, or the z axis:

$$\frac{dh}{dt} = -v\frac{\partial h}{\partial z} + Q_k \Rightarrow \frac{dh_k}{dt} = \frac{v}{\delta z}(h_{k-1} - h_k) + Q_k \tag{A.20}$$

v is the horizontal speed, h_k is the biomass enthalpy in cell k and h_{k-1} is the enthalpy in the western neighbor. Q_k represents the transfer of energy from steam to biomass in cell k. As part of this coupled partial differential equation (PDE) system, the bottom and top temperature measurements (sensors) are used to construct the

boundary conditions. The PDE model is then solved for obtaining the reactor temperature gradient, which is then utilized as the temperature of the reaction in calculating the reaction rates vector R_k.

Steam is injected through the bottom of the reactor and gets in direct contact with the biomass. The steam injection heat transfer is lumped into the boundary conditions of Equation (A.20). The biomass is assumed to heat till the steam temperature near the inlet of the reactor, and used as a western boundary condition. The heat transfer rate is computed as:

$$Q_0 = F_{in}\frac{h_f - h_0}{h_{s_0} - h_0} \tag{A.21}$$

where Q_0 is the heat transfer rate from the boundary conditions, F_{in} is the flow rate of biomass, h_f is the final enthalpy of heated biomass, h_0 is the initial biomass enthalpy, and h_{s_0} is the fresh steam enthalpy. h_0 is obtained by measuring the temperature of the biomass entering the reactor T_0:

$$h_0 = c_b(T_0 - T_r) \tag{A.22}$$

where c_b is the specific heat of biomass, and $T_r = 0$ is the reference temperature. Biomass is assumed to have a constant specific heat of approximately $c_b = 3.8\,\mathrm{kJ\,kg^{-1}}$, a value slightly lower than water ($4.18\,\mathrm{kJ\,kg^{-1}}$) since the pretreatment slurry is a mix of condensed water and biomass.

h_{s_0} is derived from saturated steam tables following the IAPWS-IF97 standard [69] and from a temperature sensor T_{s_0} mounted in the steam supply pipe:

$$h_{s_0} = f(T_{s_0}) \tag{A.23}$$

h_f is computed using the temperature of the steam layer T_s measured by the top temperature sensor, and assuming that the biomass heats till the steam temperature near the reactor inlet:

$$h_f = c_b(T_s - T_r) \tag{A.24}$$

It is natural to use the same grid in Equation (A.20) as the one from the composition balance section. The model tracks the biomass enthalpy throughout each cell of the grid. The conductive heat from the steam to the biomass layer is neglected due to the fact that biomass acts as an insulator. Therefore, only convective effects remain in the biomass layer and $Q_k = 0, 0 < k \leq N$.

The temperature profile is obtained by dividing the enthalpy from each cell with the specific heat for biomass constant c_b:

$$T_k = \frac{h_k}{c_b} \tag{A.25}$$

where T_k is the biomass temperature from cell k, and c_b is the specific heat of pretreated biomass.

A.3.5 Model Summary

The thermal reactor model tracks $n_C = 14$ species concentrations shown in vector C_k from Equation (A.18): cellulose, xylan, arabinan, lignin, acetyls, glucose, xylooligomers, xylose, arabinose, acetic acid, furfural, 5-HMF, water, and other components.

The total number of states n_x is variable depending on the initial value of N, the amount of cells in the reactor grid. n_x can be calculated as follows:

$$n_x = N \cdot (n_C + n_h) + (n_s + n_m) \tag{A.26}$$

where n_C is the number of species, $n_h = 1$ meaning one state for each grid cell enthalpy, $n_s = 2$ is the number of states from the steam layer (mass and enthalpy), and $n_m = 1$ is the total mass of biomass in the reactor. In this study, N is set to 10 leading to 153 states in total.

The model has 2 bus inputs: one for feedstock, and another one for steam. The feedstock input has 16 components: flow rate (1), feedstock concentrations (14), and enthalpy (1). The steam input has 2 components: flow rate (1), and enthalpy (1). In total there are 18 inputs.

There are 2 bus outputs: pretreated fibers and the liquid rich in C5 sugars. Each bus has 16 components: flow rate (1), species concentrations (14), and enthalpy (1).

Table A.3 shows the fixed model parameters. The kinetics parameters are determined in the model calibration section of this study.

A.4 Results and Discussion

This section is split into a model analysis and validation part, and the model application as a state estimator. The model analysis and validation section contains the sensitivity and uncertainty analysis, parameter estimation, and residual analysis.

A.4.1 Model Analysis and Validation

A.4.1.1 Model Initialization

Model parameters are calibrated with respect to the following NIR measurements: cellulose, xylan, lignin, acetic acid and furfural. The data were obtained from a demonstration scale thermal reactor throughout 15 h of operation. Only a subset of 7 h is used for calibration and parameter estimation, while the entire set of measurements is used for validation and residual analysis. The measurements of cellulose, xylan and lignin are reported as percentage of dry matter, while acetic

Table A.3: Fixed model parameters. Feedstock is soaked before entering the thermal reactor till approximately 40 % dry matter.

Parameter	Description	Value	Unit
L_r	Thermal reactor length	12	m
t_r	Pretreatment retention time	15	min
N	Grid cell resolution	10	-
δz	Grid cell width	1.2	m
v	Reactor horizontal speed	0.013	$\mathrm{m\,s^{-1}}$
F_{in}	Feedstock flow rate	6	$\mathrm{kg\,s^{-1}}$
h_0	Feedstock enthalpy	117	$\mathrm{kJ\,kg^{-1}}$
h_{s_0}	Fresh steam enthalpy (saturated)	2795	$\mathrm{kJ\,kg^{-1}}$
c_b	Specific heat of pretreated biomass (constant)	3.8	$\mathrm{kJ/(kgK)}$
R	Universal gas constant	8.3145	$\mathrm{J/(molK)}$
C_0	Feedstock composition:		
	Cellulose	160	$\mathrm{g\,kg^{-1}}$
	Xylan	95	$\mathrm{g\,kg^{-1}}$
	Arabinan	8	$\mathrm{g\,kg^{-1}}$
	Lignin	80	$\mathrm{g\,kg^{-1}}$
	Acetyls	16	$\mathrm{g\,kg^{-1}}$
	Glucose	0	$\mathrm{g\,kg^{-1}}$
	Xylooligomers	0	$\mathrm{g\,kg^{-1}}$
	Xylose	0	$\mathrm{g\,kg^{-1}}$
	Arabinose	0	$\mathrm{g\,kg^{-1}}$
	Acetic acid	0	$\mathrm{g\,kg^{-1}}$
	Furfural	0	$\mathrm{g\,kg^{-1}}$
	5-HMF	0	$\mathrm{g\,kg^{-1}}$
	Water	600	$\mathrm{g\,kg^{-1}}$
	Other	41	$\mathrm{g\,kg^{-1}}$
		1000	$\mathrm{g\,kg^{-1}}$

acid and furfural concentrations as g/kg of slurry. Table A.4 illustrates the estimated values while Figure A.5 shows the fitting result.

Table A.4: Calibrated parameters using the estimation dataset from Figure A.2, which provides system dynamics in the 178 °C to 185 °C range.

Parameter	Value	Units
k_{Xo}	2.78×10^{31}	s^{-1}
E_{Xo}	298 000	$J\,mol^{-1}$
k_X	1.31×10^{34}	s^{-1}
E_X	305 000	$J\,mol^{-1}$
k_G	1.11×10^{35}	s^{-1}
E_G	336 000	$J\,mol^{-1}$
k_{PL}	1.03×10^{33}	s^{-1}
E_{PL}	326 000	$J\,mol^{-1}$
k_F	5.09×10^{33}	s^{-1}
E_F	327 000	$J\,mol^{-1}$
k_{Ac}	4.88×10^{24}	s^{-1}
E_{Ac}	243 000	$J\,mol^{-1}$
k_H	1×10^{31}	s^{-1}
E_H	300 000	$J\,mol^{-1}$
a_{PL}	0.102	–

A.4.1.2 Sensitivity Analysis

The sensitivity analysis contributes to a good process understanding by quantifying the relation between outputs and model parameters. The analysis is carried with respect to all model parameters from Table A.4 and aims at ranking these parameters by their significance. In this way a reduced set of relevant parameters can be identified and used for a proper parameter estimation procedure. Also, by fitting less parameters, the model calibration procedure simplifies and can be run more often in a real industrial application.

A measure of sensitivity with respect to model parameters, and suitable for time varying signals, is the delta mean square δ_{ik}^{msqr} defined by Brun et al. [64]:

$$\delta_{ik}^{msqr} = \sqrt{\frac{1}{N} s_{nd,ik}^{\top} s_{nd,ik}} \tag{A.27}$$

where k is the parameter index, i is the model output index, N is the number of samples, and $s_{nd,ik}$ is a vector with the non dimensional sensitivity calculated in each

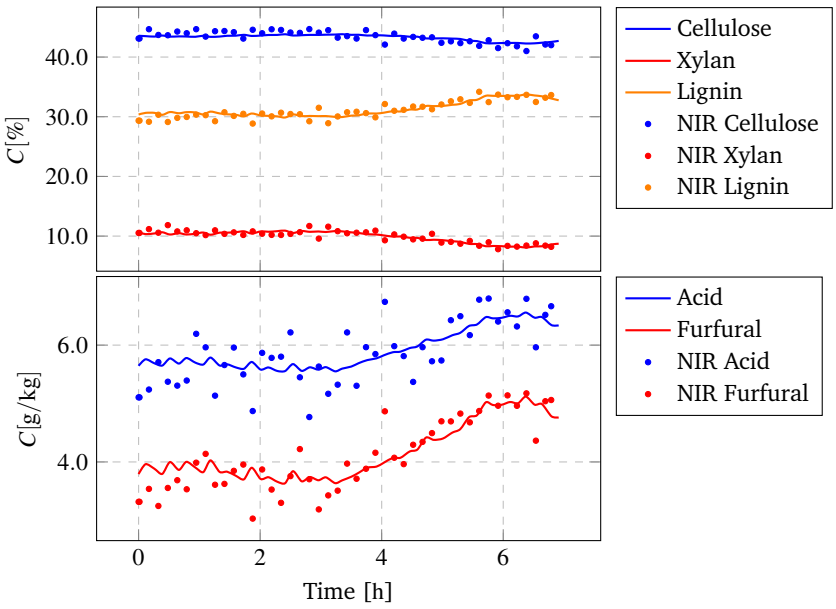


Figure A.5: Solid and liquid content of pretreated biomass: cellulose, xylan, lignin, acetic acid, and furfural.

sample:

$$s_{nd,ik} = \frac{\partial y_i}{\partial \theta_k} \frac{\theta_k}{sc_i} \tag{A.28}$$

$\partial y_i / \partial \theta_k$ represents the output variation with respect to parameter θ_k, and sc_i is a scaling factor with the same physical dimension as the corresponding observation in order to make this measure non dimensional. In this study, the scaling factor is chosen as the mean value of output i:

$$sc_i = \frac{1}{N} \sum_{1}^{N} y_i(k) \tag{A.29}$$

All parameters are ranked according to δ_{ik}^{msqr} for each output i. As the sensitivity measure is non-dimensional, a cumulative variable is also defined as the sum of sensitivities for a given parameter in all outputs. Because the model has to predict all defined outputs, the subset of significant parameters contains all parameters with a cumulative sensitivity above a threshold, which is set to 2 % of the maximum sensitivity. The cumulative delta mean square is defined as:

$$\delta_k^{msqr} = \sum_{i=1}^{n_y} \delta_{ik}^{msqr} \tag{A.30}$$

where n_y is the total number of outputs, i.e. $n_y = 5$ in this study: concentrations of cellulose, xylan, lignin, acetic acid, and furfural.

The same sensitivity analysis methodology has been applied on a cellulosic hydrolysis model in previous studies by Sin et al. [8], and Prunescu and Sin [6].

The sensitivity analysis results for the pretreatment process can be observed in Figure A.6:

- Cellulose is mostly sensitive to E_G, which is expected since cellulose is directly hydrolyzed into glucose and this reaction is sensitive to the reactor temperature.
- Xylan is sensitive to the activation energy of xylooligomers production E_{Xo}, which is not surprising since xylooligomers are direct products of xylan hydrolysis.
- Lignin as percentage of dry matter follows the changes from xylan and cellulose content, which is the solid content of biomass. This means that if more xylan is hydrolyzed then the percentage of lignin in the remaining slurry after separation will increase. There is also lignin production as pseudo-lignin and E_{PL} is ranked second. E_F, E_G and E_X influence the amount of carbohydrates that participate in pseudo-lignin formation.
- Organic acids, mostly represented by acetic acid, is sensitive to the activation energy for reaction rate r_{Ac}, i.e. E_{Ac}.
- The last output, furfural, is mostly sensitive to the activation energy E_F of reaction rate r_F. E_{Xo} appears second because it directly affects the amount of xylose, which degrades to furfural.

The cumulative sensitivity measure is illustrated in the right bottom plot of Figure A.6. The most sensitive parameters are picked to be the first six:

$$\theta_R = [E_{Xo} \quad E_F \quad E_{Ac} \quad E_G \quad E_{PL} \quad E_H] \tag{A.31}$$

which are all activation energies directly involved in the reaction temperature dependency. This is natural as it has been observed in experimental studies that small changes in reactor temperature impact significantly the composition of pretreated fibers, and is in agreement with process expert knowledge [50]. This analysis also ranks the activation energies among themselves. E_{Xo} (related to xylooligomers production) is ranked first as the most sensitive parameter, which tells that hemicellulose hydrolysis is the main phenomenon taking place in the reactor. E_F is ranked second showing that furfural is the main by-product followed by acetic acid (E_{Ac}). E_G is ranked 4th, which means that cellulose hydrolysis also occurs in the reactor but at a much lower rate than hemicellulose hydrolysis. The other two by-products, i.e. pseudo-lignin and 5-HMF, have a lower significance.

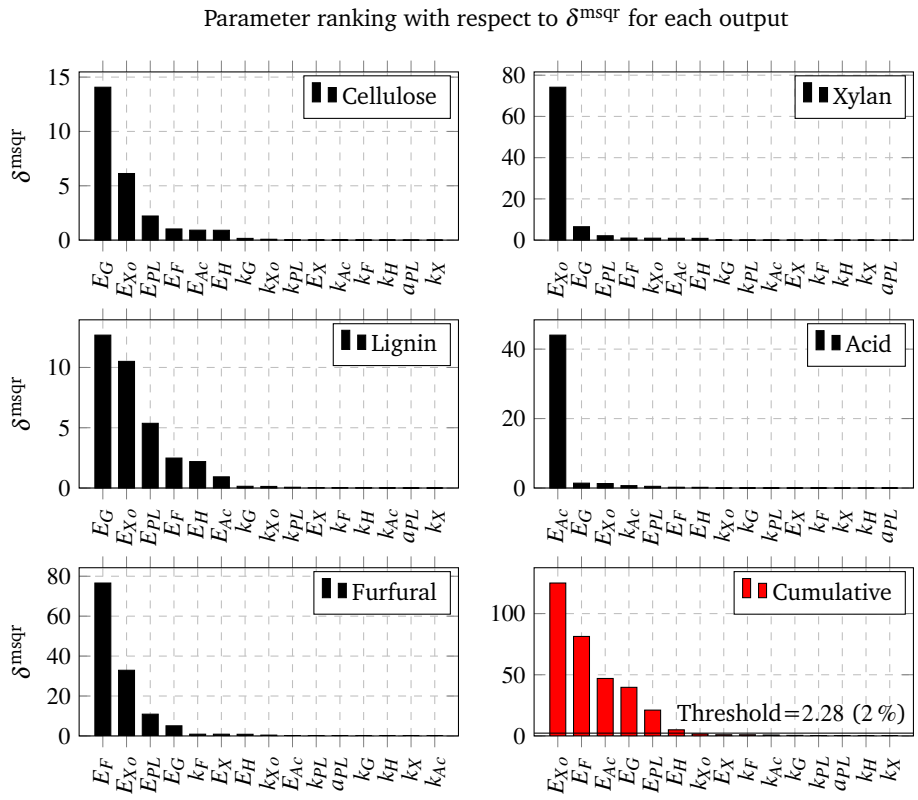


Figure A.6: Sensitivity analysis by calculating the δ^{msqr} for each output with respect to each model parameter. Although the model has 15 parameters in total, only 6 parameters are of interest as indicated in the bottom right figure.

A.4.1.3 Parameter Estimation

The reduced set of parameters θ_R is identified based on the real NIR measurements from the demonstration scale facility. A nonlinear least square method is run to obtain the parameter estimates $\hat{\theta}_R$ along with their standard deviation σ and correlation matrix $\mathbf{R}_\theta$. Table A.5 shows the results. The estimated parameter values are deemed credible as the parameter estimation error indicated by the standard deviation is rather low, i.e. less than 1 %. However some of the parameter estimates are found to be significantly correlated, e.g. correlation between E_G and E_{Xo} is 0.74, which implies poor identifiability. The reason for this is the dataset used for parameter estimation. The data are obtained from an industrial scale facility during normal operational conditions under small temperature disturbances. Such data with limited dynamics cannot be expected to provide rich information for complete identification of all the parameters [98] and design of experiments for identification

should be pursued on lower scale facilities. Other measurements should also be included in the parameter estimation analysis, such as xylooligomers, xylose and glucose content of the liquid part, which are missing in this study.

Table A.5: Parameter estimation results. The estimates of the most important parameters with their standard deviations are shown in Table A.6(a). The correlation matrix is displayed in Table A.6(b). The results are valid in the 178 °C to 185 °C range.

(a) Parameter estimates in $\mathrm{J\,mol^{-1}}$ with standard deviation σ and 95 % confidence interval.

Parameter	Value [$\mathrm{J\,mol^{-1}}$]	Standard deviation σ	Lower bound	Upper bound
E_{Xo}	298 010	98	297 918	298 102
E_F	327 255	285	326 989	327 520
E_{Ac}	242 693	174	242 531	242 855
E_G	335 616	249	335 383	335 848
E_{PL}	325 632	1573	324 165	327 099
E_H	299 999	2639	297 537	302 461

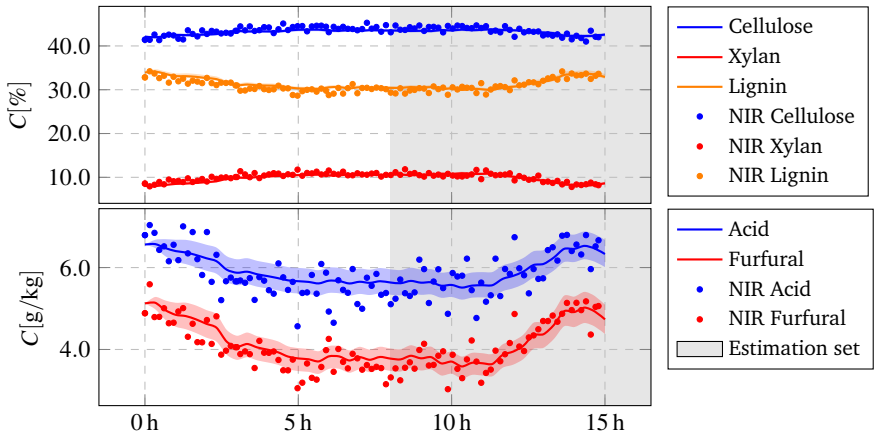
(b) Correlation matrix $\mathbf{R}_\theta$.

Parameter	E_{Xo}	E_F	E_{Ac}	E_G	E_{PL}	E_H
E_{Xo}	1					
E_F	−0.51	1				
E_{Ac}	0.17	−0.12	1			
E_G	0.74	−0.54	0.26	1		
E_{PL}	−0.51	0.74	−0.15	−0.63	1	
E_H	0.14	−0.61	0.01	0.16	−0.85	1

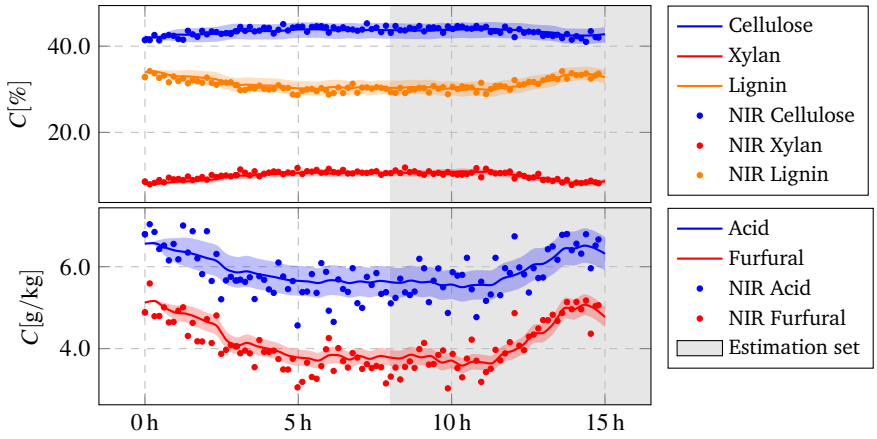
A.4.1.4 Uncertainty Analysis

In order to assess the uncertainty of the model, a large number of Monte Carlo simulations are run in 3 different scenarios that cover most sources of uncertainty: model parameters, feed parameters, and combined.

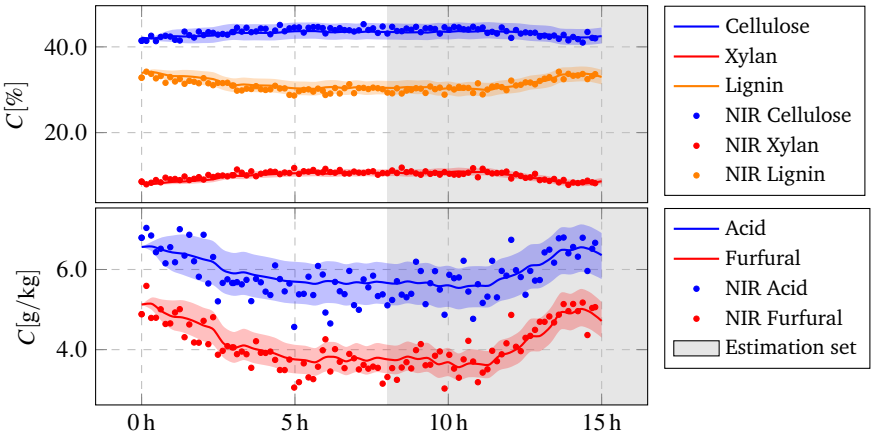
1. Model parameters uncertainty: the standard deviation and the correlation matrix for model parameters were obtained in the previous step. A number of 200 samples are generated using Latin hypercube sampling with correlation control. Figure A.7(a) shows the biomass composition predictions with the 5th-95th percentile interval considering only model parameters uncertainty. The model parameters uncertainty cannot explain the entire output variations, especially in the solid content of the mixture. This is expected since there



(a) Final yield uncertainty with respect to model parameters.



(b) Final yield uncertainty with respect to feed parameters.



(c) Final yield uncertainty with respect to model and feed parameters (combined).

Figure A.7: Model predictions uncertainty.

are other sources of uncertainty, i.e. initial biomass composition or feed parameters.

2. Feed parameters uncertainty: Latin hypercube uniform sampling is performed to cover 7 % variation of the initial composition and another 200 samples are generated considering fixed model parameters. The simulation results with NIR measurements and percentile intervals are displayed in Figure A.7(b). Cellulose and lignin concentrations are well within the bounds while xylan exceeds the interval in some samples. Acetic acid and furfural are not entirely captured by the percentile interval, but it might be due to model parameters uncertainty.

3. Combined uncertainty: the results are shown in Figure A.7(c). The uncertainty for acetic acid and furfural is wider but still not entirely within bounds. NIR instruments have a lower accuracy for low concentrations and it is expected that the remaining prediction inaccuracy is due to non Gaussian measurement error. Acetic acid and furfural have a much lower concentration compared to the solids, i.e. cellulose, xylan and lignin. Therefore, it is expected to have significant measurement errors at these low concentrations.

A.4.1.5 Reactor Profiles

The thermal reactor was split in 10 smaller cells and reactor profiles can be drawn in order to observer how biomass changes composition along the reactor. The top plot from Figure A.8 indicates the reaction temperature, which is not constant but rather a gradient built with the help of the top and bottom temperature sensors. The biomass is subject to a range of pretreatment temperatures. The remaining 3 plots from Figure A.8 show how biomass composition changes along the reactor. The figure x axis is the length of the reactor, while the y axis contains the species concentration in $\mathrm{g/kg}$. The second plot indicates that arabinan is fully hydrolyzed, while cellulose and xylan are partially hydrolyzed with more xylan conversion than cellulose. The following plot shows sugar production, i.e. glucose, xylooligomers, xylose and arabinose. By-products such as acetic acid and furfural are displayed in the bottom plot of Figure A.8.

A.4.1.6 Standardized Regression Coefficients (SRC)

A linear model is first fitted from model parameters to each model output from the Monte Carlo simulations [66, 67]:

$$y_{reg} = a + \sum_k b_k \theta_k \tag{A.32}$$

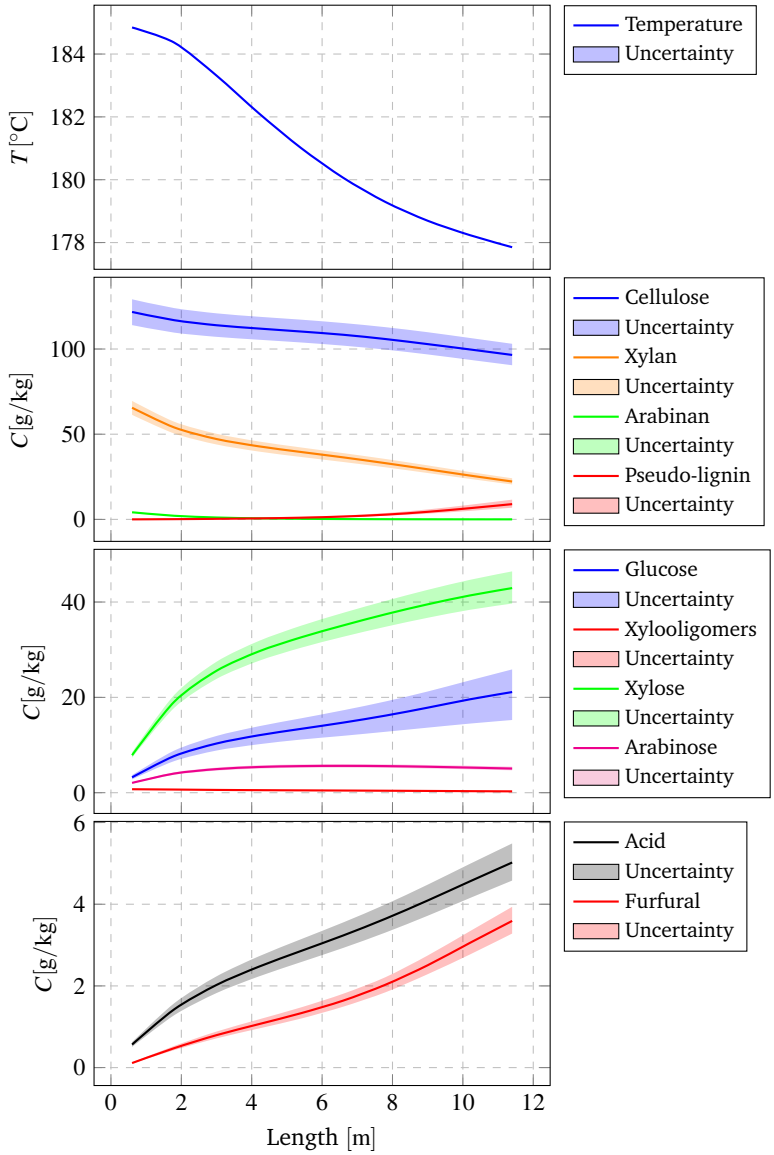


Figure A.8: The top plot shows the reactor horizontal temperature gradient, which is used for calculating the reaction rates. The other plots illustrate the reactor conversion profiles with confidence bounds due to both model and feed parameters uncertainty.

where y_{reg} is the i^{th} output, and a and b_k are the linear model parameters. The standardized regression coefficients β are a global sensitivity measure and are defined as:

$$\beta_k = \frac{\sigma_{\hat{\theta}_{R_k}}}{\sigma_{y_i}} b_k \tag{A.33}$$

where β_k is the β coefficient, $\sigma_{\hat{\theta}_{R_k}}$ is the standard deviation of the parameter estimate, σ_{y_i} is the standard deviation of output i, and b_k is the linear model parameter. β_k is an indicator for how much the parameter uncertainty contributes to the prediction uncertainty.

The β coefficients are displayed in Table A.6 and explained below:

Table A.6: SRC coefficients for model and feed (combined) parameters.

θ	Cellulose	θ	Xylan	θ	Lignin	θ	Acetic Acid	θ	Furfural
C_{C_S}	0.82	C_{X_S}	0.70	C_{L_S}	0.77	C_{AC_S}	0.72	E_F	−1.00
C_{L_S}	−0.43	E_{Xo}	0.54	E_{PL}	−0.72	E_{Ac}	−0.60	E_{PL}	0.75
E_G	0.38	C_{C_S}	−0.38	C_{C_S}	−0.47	C_{C_S}	0.06	C_{X_S}	0.56
E_{PL}	0.36	C_{L_S}	−0.23	E_H	−0.47	C_{L_S}	0.04	E_{Xo}	−0.16
E_H	0.23	E_{PL}	0.20	E_G	−0.24	E_{PL}	−0.04	C_{A_S}	0.08
C_{X_S}	−0.23	E_H	0.14	E_{Xo}	−0.11	E_H	−0.02	E_H	0.07
E_{Xo}	−0.08	E_G	−0.12	E_F	−0.06	E_G	0.02	C_{L_S}	0.04
C_{AC_S}	−0.05	C_{AC_S}	−0.03	C_{X_S}	−0.04	C_{X_S}	0.01	C_{C_S}	0.03
E_F	0.02	E_F	0.02	C_{AC_S}	−0.04	E_{Xo}	0.01	E_G	0.02
E_{Ac}	−0.02	E_{Ac}	−0.02	E_{Ac}	−0.02	E_F	−0.00	C_{AC_S}	0.01
C_{A_S}	−0.01	C_{A_S}	−0.01	C_{A_S}	0.02	C_{A_S}	−0.00	E_{Ac}	0.00
R^2	1.00	R^2	1.00	R^2	0.99	R^2	1.00	R^2	0.93

- Concentration of solid cellulose C_{C_S}: is sensitive to the initial concentration of cellulose and lignin from feedstock, to the activation energy of glucose reaction r_G, and pseudo-lignin reaction r_L. Lignin appears due to the fact that the function of lignin in nature is to protect cellulose from degradation, thus a larger amount of lignin necessarily affects the kinetics of cellulose degradation by reducing it (the corresponding β coefficient has a negative value). There is also a percentage redistribution of biomass content when lignin or pseudo-lignin increase. Thus, production of pseudo-lignin causes a decrease in cellulose concentration in pretreated fibers and E_{PL} appears among the top parameters.

- Concentration of solid xylan C_{X_S}: is sensitive to the initial concentration of xylan in feedstock, and to xylan hydrolysis or the activation energy of xy-

looligomers reaction rate E_{Xo}. Higher concentrations of cellulose and lignin in feedstock can cause a decrease in xylan content due to percentage redistribution.

- Concentration of solid lignin C_{L_S}: high sensitivity is detected in initial lignin concentration from feedstock and in pseudo-lignin production. Initial concentration of cellulose also has an impact due to percentage redistribution of fiber content.
- Concentration of acid C_{Ac}: as expected, acetyls content in feedstock and activation energy of the acid production reaction rate E_{Ac} have the greatest impact on acid content in pretreated fibers.
- Concentration of furfural C_F: activation energy of the furfural production reaction E_F and pseudo-lignin E_{PL} are the most sensitive parameters. Furfural participates in pseudo-lignin formation and is expected to find E_{PL} among the top parameters.

Parameters related to feed composition have a higher sensitivity than the kinetic parameters, even though only a deviation of 7 % was introduced in the initial biomass composition. This indicates the importance of measuring the initial composition of feedstock for more accurate model predictions. The SRC based sensitivity results are credible as the degree of linearization indicated by Pearson correlation coefficient R^2 is high for all the outputs [66].

A.4.1.7 Residual Analysis

Statistical analysis of residuals is important for model validation especially as it provides information on any model bias in the predictions [99]. The prediction estimation error or residual is calculated as:

$$e = y - \hat{y} \tag{A.34}$$

where e is the residual vector, y is the real measurement and $\hat{y}$ is the estimated output. There are 5 measurements regarding biomass composition, i.e. cellulose (%), xylan (%), lignin (%, includes pseudo-lignin), acetic acid (g/kg) and furfural content (g/kg). The residuals are calculated based on the validation set and they are displayed in the top plot series from Figure A.9. The estimation errors for cellulose, xylan and lignin stay within -2 to 2 %, while acetic acid and furfural are slightly lower, -1 to 1 g/kg.

If the residuals are white noise (i.e. Gaussian distribution with mean zero and a certain standard deviation) and uncorrelated then there is no systematic error or

bias in the model. To check the hypothesis of residual Gaussian distribution, a 5 % t-test is performed. The test is passed and the Gaussian probability plots from Figure A.9 (the middle series) show how close the residuals are to a Normal distribution. The autocorrelation determines if there still is any information in the residuals that is not captured by the model. The bottom series of plots from Figure A.9 display the autocorrelation function at 50 lags with the 95 % confidence interval. There are spikes in almost all plots except acetic acid production that exceed the confidence interval but they are not significant. The feedstock composition and the assumption that it remains constant might be a good source for these results. In reality feedstock has composition variations.

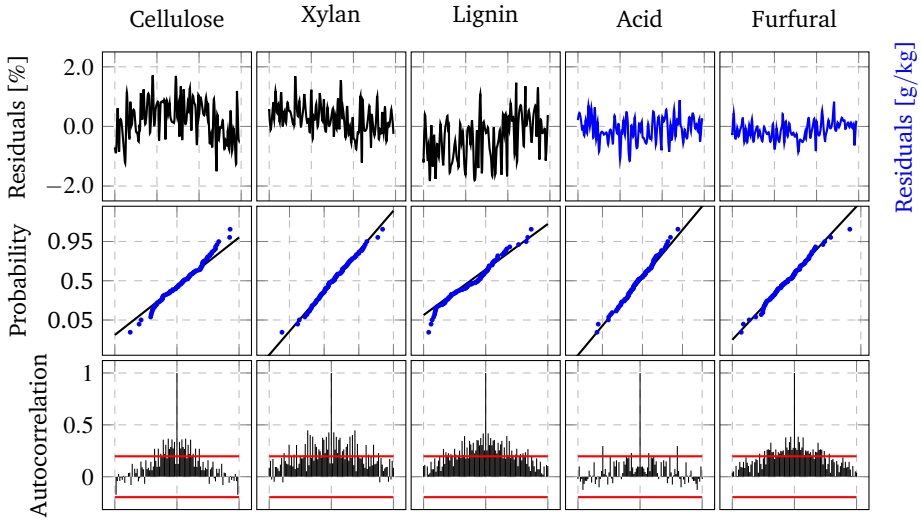


Figure A.9: Residual analysis for the validation set. The top plot series show the residuals; the middle series compare in a Gaussian probability plot the distribution of residuals to a Normal distribution; the bottom series display the autocorrelation function and the 95 % confidence interval.

A.4.2 Model Application: State Estimator

The utility of the model for industrial application is highlighted through the development of a state estimator using a Kalman filter. The state estimator is useful for performance monitoring by estimating in real-time the entire biomass composition, i.e. both sugars and inhibitors production. The estimator infers all species concentrations from pretreated fibers using a reduced number of measurements provided by the NIR instrument.

A state estimator is naturally based on the nonlinear process model:

$$\begin{cases} \dot{\hat{x}} = f(\hat{x}, u) \\ \hat{y} = g(\hat{x}) \end{cases} \tag{A.35}$$

where $\hat{x}$ is the state estimates vector, $\hat{y}$ is the output estimates vector, $f(\hat{x}, u)$ is the nonlinear state derivatives as a function of states and inputs, and $g(\hat{x})$ is the nonlinear output function.

Several methods exist for use of nonlinear process models in state estimation. Classical approaches include an extended Kalman filter for combined state and parameter estimation [100] for a linearized system at a particular point of operation, or direct inclusion of the nonlinear process model in the filter [101]. Later developments have included the unscented Kalman and particle filters to better explore and approximate non-Gaussian nature of the process noise in a nonlinear system [102].

To compensate for model-mismatches between estimates and real measurements, an extra correction term is added to the state derivatives equation from (A.35):

$$\begin{cases} \dot{\hat{x}} = f(\hat{x}, u) + \mathbf{L}e \\ \hat{y} = g(\hat{x}) \end{cases} \tag{A.36}$$

where e is the estimation error defined in Equation (A.34), and $\mathbf{L}$ is a gain matrix that needs to be designed. The expression in (A.36) is known as a high-gain observer. The block diagram of the state estimator for the pretreatment process is shown in Figure A.10. In this application, feedback from the NIR instrument is used to create the estimation error term. A constant feedstock composition is assumed and it can be determined in reality by analyzing stock samples offline in the laboratory.

There are various methods to calculate the gain matrix $\mathbf{L}$ among which the most common ones include pole-placement and extended Kalman filter (EKF). Pole-placement and extended Kalman filters rely on a linearized version of the process model. The extended Kalman filter uses the linearized model to calculate the error covariance matrix and, from this, the observer gain. If the linearized model and process noise covariance matrices are fixed then the error covariance matrix and the observer gain can be calculated offline. Kalman filters are known to be optimal in the sense of minimizing the estimation error covariance matrix in the presence of process and measurement noise [103], and are preferred to pole-placement techniques. Kalman filters require known process noise covariance matrices to work optimally, which is rather difficult in reality to approximate but alternatives exist, which estimate the noise structure online [100], at the expense of more complexity of the estimation algorithm.

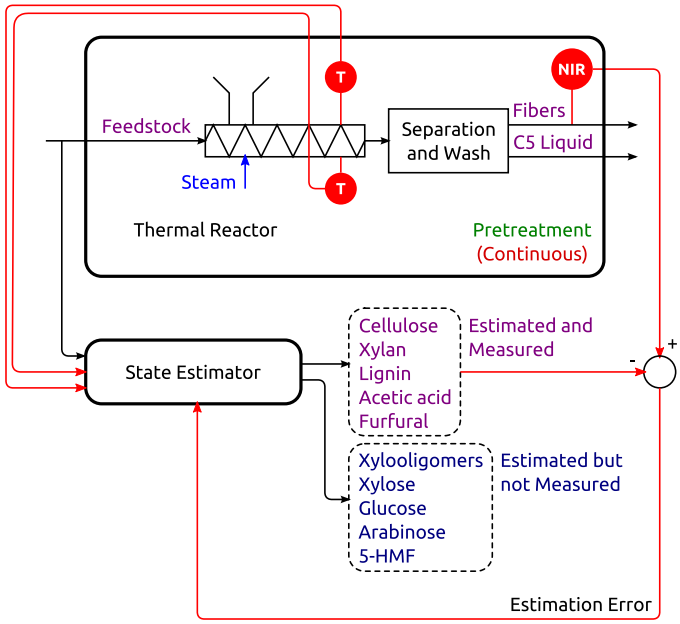


Figure A.10: State estimator block diagram. The state estimator uses 2 temperature sensors and the NIR measurements to infer pretreated biomass composition. The state estimator acts both as measurement filter and soft sensor.

The operational point in this study does not change significantly and this is the reason why a static extended Kalman filter is chosen for the state estimator. The EKF design process follow these steps:

1. The first step is to obtain a stochastic linear model by linearizing the nonlinear process model from (A.35) around the nominal operational point seen in the datasets:

$$\begin{cases} \dot{\hat{x}} = \mathbf{A}\hat{x} + \mathbf{B}u + \mathbf{G}w \\ \hat{y} = \mathbf{C}\hat{x} + v \end{cases} \tag{A.37}$$

where $\mathbf{A}$ is the dynamic matrix of the linearized model, $\mathbf{B}$ is the input matrix, $\mathbf{G}$ is the state noise propagation matrix, and $\mathbf{C}$ is the output matrix. State noise w and measurement noise v are assumed to be 0 mean uncorrelated white noise sequences with variances $\mathbf{Q}$ and $\mathbf{R}$:

$$w \sim (0, \mathbf{Q}) \qquad v \sim (0, \mathbf{R}) \tag{A.38}$$

The linear model matrices are calculated by differentiating the nonlinear

process model around the nominal operational point (x_e, u_e):

$$\mathbf{A} = \left.\frac{\partial f(\hat{x},u)}{\partial \hat{x}}\right|_{\hat{x}=x_e,u=u_e} \qquad \mathbf{B} = \left.\frac{\partial f(\hat{x},u)}{\partial u}\right|_{\hat{x}=x_e,u=u_e} \qquad \mathbf{G} = \left.\frac{\partial f(\hat{x},u)}{\partial w}\right|_{\hat{x}=x_e,u=u_e}$$
$$\mathbf{C} = \left.\frac{\partial g(\hat{x})}{\partial \hat{x}}\right|_{\hat{x}=x_e} \tag{A.39}$$

where x_e and u_e form the nominal operational point in terms of states and inputs. It is not known how the process noise propagates inside the system dynamics, and $\mathbf{G}$ is set to $\mathbf{I}_{n_x}$ (identity matrix of size n_x or total number of states).

2. The second step is to approximate the state and measurement noise covariance matrices, i.e. $\mathbf{Q}$ and $\mathbf{R}$, which are set to:

$$\mathbf{Q} = 10^{-4} \cdot \mathbf{I}_{n_x} \qquad \mathbf{R} = \begin{bmatrix} 50 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 50 & 0 & 0 \\ 0 & 0 & 0 & 600 & 0 \\ 0 & 0 & 0 & 0 & 100 \end{bmatrix} \tag{A.40}$$

where $\mathbf{I}_{n_x}$ is an identity matrix, and n_x is the number of states. The concentrations of solids are higher and more reliable, therefore lower variances are used in the first 3 diagonal terms from $\mathbf{R}$, which correspond to cellulose, xylan and lignin (measured in % of dry matter). The other 2 diagonal numbers are the variances for acetic acid and furfural (measured in $\mathrm{g\,kg^{-1}}$), which are in low concentrations and have larger measurement errors.

3. In the last step of the design process, the static Kalman gain is calculated [103]:

$$\mathbf{L} = \mathbf{P}\mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1} \tag{A.41}$$

where $\mathbf{P}$ is the error covariance matrix found from solving the Riccati equation [103]:

$$\dot{\mathbf{P}} = \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^{\mathrm{T}} + \mathbf{G}\mathbf{Q}\mathbf{G}^{\mathrm{T}} - \mathbf{P}\mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{C}\mathbf{P} \tag{A.42}$$

when $\dot{\mathbf{P}} = 0$.

The Kalman state estimator is tested throughout the whole dataset of 15 h. Figure A.11 shows the model outputs overlapped with the NIR measurements. The model and the Kalman filter succeed in following the dynamic trends of the process. The

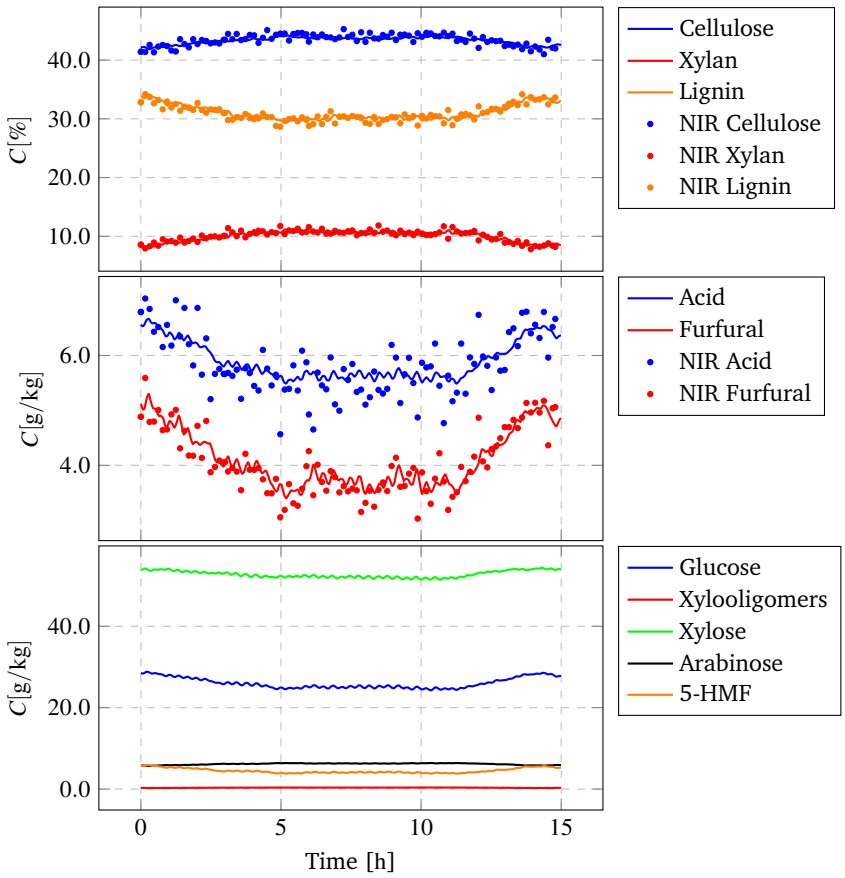


Figure A.11: Pretreatment state estimator throughout 15 h of operation. The plots show the solid and liquid content of pretreated biomass: cellulose, xylan, lignin in the top plot, acetic acid and furfural in the second plot, and estimations of by-products production in the bottom plot.

state estimator filters the NIR measurements and also acts as a soft sensor for by-products production: glucose, xylooligomers, xylose, arabinose and 5-HMF shown in the bottom plot of Figure A.11.

The residuals are displayed in Figure A.12. The variance of the raw residuals is similar to the one from Figure A.9. However, the Kalman filter is able to capture more information from the process causing the autocorrelation function to stay within the confidence interval (the bottom plot series from Figure A.12). The residuals distribution slightly change as indicated by the middle series of plots from Figure A.12. This happens because the noise covariance matrices, and the propagation of noise through the system are unknown and set to arbitrary values.

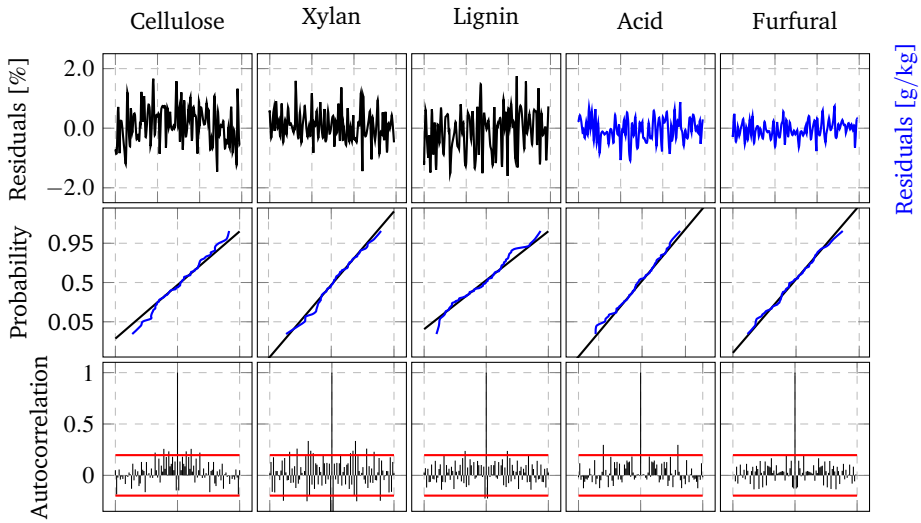


Figure A.12: Residual analysis for the validation set with estimator. The top plot series show the residuals; the middle series compare in a Gaussian probability plot the distribution of residuals to a Normal distribution; the bottom series display the autocorrelation function and the 95 % confidence region.

The state estimator successfully embeds the real measurements to compensate on model predictions mismatches under uncertainties such variation in feedstock composition or in model parameters. Also, the state estimator acts as a soft sensor for several unmeasured variables, some of which act as inhibitors in downstream processing especially in enzymatic hydrolysis and fermentation. Such information includes: xylose, xylooligomers, arabinose, pseudo-lignin, glucose, and 5-HMF production.

A.5 Conclusions

This study presented a dynamic model for a large scale biomass hydrothermal pretreatment process. The model was capable of predicting the composition of pretreated fibers, and has been properly analyzed assessing its sensitivity and uncertainty taking into account both model and feed parameters. The sensitivity analysis showed that the set of 15 model parameters can be reduced to a subset of 6 significant parameters that deal with the activation energy of most reactions. This finding was expected as the activation energy appears in the temperature dependency term of the reaction equation, and small reactor temperatures deviations can lead to significant changes in pretreated biomass. Real data were extracted from a demonstration

scale biorefinery and used for parameter estimation and model validation. The data exhibit limited system dynamics and provide a range of validity for the parameter estimates between 178 °C to 185 °C. In order to extend the validity range of the parameter estimates to different temperature ranges, additional experimental data should be collected either by experimenting at lower scale facilities or changing the operational points of the biorefinery, which is not easily accessible.

The uncertainty analysis was conducted with Latin hypercube sampling of both model and feed parameters, and Monte Carlo simulations. The 5th-95th percentile includes most of the real data variations, except the very low concentrations of acetic acid and furfural, which is expected as the NIR instrument is affected by high measurement errors at low concentrations. The global sensitivity analysis showed that feedstock composition parameters have a higher significance than model parameters, suggesting that feedstock composition should be measured upstream for more accurate predictions.

In the last section of the study, a Kalman state estimator was designed as an application of the process model. The estimator was then tested on the whole dataset, i.e. throughout 15 h of operation. The model predictions were found to follow the dynamic trends of the process, making it valuable for control design, diagnosis, real-time optimization, and inferential sensors development at full scale applications. The state estimator acts both as a measurement filter and soft sensor for estimating unmeasured variables, i.e. C5 sugars production and inhibitors, which are important in downstream processes.

Paper B

Dynamic Modeling and Validation of a Lignocellulosic Enzymatic Hydrolysis Process - A Demonstration Scale Study

Remus Mihail Prunescu[1], Gürkan Sin[*2]

[1]Department of Electrical Engineering, Automation and Control Group, Technical University of Denmark, Elektrovej Building 326, 2800, Kgs. Lyngby, Denmark

[2]CAPEC-PROCESS, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads Buildings 227 and 229, 2800, Kgs. Lyngby, Denmark

Abstract:
The enzymatic hydrolysis process is one of the key steps in second generation biofuel production. After being thermally pretreated, the lignocellulosic material is liquefied by enzymes prior to fermentation. The scope of this paper is to evaluate a dynamic model of the hydrolysis process on a demonstration scale reactor. The following novel features are included: the application of the Convection-Diffusion-Reaction equation to a hydrolysis reactor to assess transport and mixing effects; the extension of a competitive kinetic model with enzymatic pH dependency and hemicellulose hydrolysis; a comprehensive pH model; and viscosity estimations during the course of reaction. The model is evaluated against real data extracted from a demonstration scale biorefinery throughout several days of operation. All measurements are within predictions uncertainty and, therefore, the model constitutes a valuable tool to support process optimization, performance monitoring, diagnosis and process control at full-scale studies.

[*]Principal corresponding author. Tel.: +45 45252806; E-mail: gsi@kt.dtu.dk

B.1 Introduction

Bioethanol is thought to become a viable alternative to fossil fuels [1] and many countries with agricultural resources show an increasing interest in the second generation biofuel production technology, especially in USA, Brazil, Denmark and Italy. The latest developments that approach biorefineries to a commercial reality are presented by Larsen et al. [3].

In biorefinery concepts that employ a biochemical conversion route, the lignocellulosic material originating from agricultural wastes is typically transformed into bioethanol following 4 major steps, i.e. pretreatment, enzymatic hydrolysis, fermentation and separation [4]. The pretreated lignocellulosic material has a high content of cellulose, hemicellulose and lignin. The purpose of the enzymatic hydrolysis process is to break down the long polymeric chains of cellulose and xylan into its constituent sugar units, mostly glucose and xylose.

The underlying mechanism of the enzymatic hydrolysis process is a complex one, which has been the target of many modeling studies. Kadam et al. [60] formulate a mechanistic conversion model while Zhang and Lynd [104] explain in detail how the cellulosic depolymerization occurs under the effect of enzymes. The Kadam model has been validated in laboratory scale experiments both by Kadam et al. [60] and Hodge et al. [105]. Sin et al. [8] have performed an identifiability and uncertainty analysis of the Kadam model and Morales-Rodriguez et al. [96] have integrated it into a dynamic modeling framework for assessing biorefinery configurations. The depolymerization model has been extended and validated in laboratory scale experiments by Hosseini and Shah [106, 107]. So far, none of these models have been evaluated at a realistic demonstration scale.

In order to formulate a model for large scale biorefineries, several extensions need to be made to the previous models. pH calculations were not performed before although it is well known that enzymes are sensitive to pH following a Gaussian curve. At laboratory scale, perfect pH control can be easily achieved and assumed but at larger scales such control becomes challenging. When the process runs in a continuous manner, the inflow composition varies due to different pretreatment conditions or biomass composition and the concentration of acetic acid has disturbances that affect the pH level [12]. Also, previous models do not include xylan hydrolysis, which contributes to acetic acid formation inside the reactor leading to a pH profile along the container that affects the enzymatic activity. Another missing feature is viscosity calculations, which are important for estimating diffusion coefficients and for assessing transportation and mixing effects.

With these in background, a dynamic lignocellulosic hydrolysis model adequate

for demonstration and large scale processes is formulated with the following features: the reactor is modeled as plug flow using the Convection-Diffusion-Reaction (CDR) equation in one dimension in order to capture properly transport and mixing effects along the length of the tank, i.e. the x axis; the reaction kinetics comprises competitive cellulose and xylan hydrolysis with temperature and pH dependency plus furfural inhibition; pH dynamics account for most of the relevant buffers that come from the pretreatment process, i.e. acetic, succinic and lactic acids plus a bicarbonate formation system with CO_2 stripping effects. Chemical kinetics, and pH and viscosity calculations are tracked in each cell of the 1-D CDR model.

The model is then analyzed in order to assess its sensitivity to kinetic and feed parameters, and to quantify its predictions uncertainty.

This paper is structured as follows: section B.2 describes a typical demonstration scale biorefinery and how measurements were recorded for comparing against model predictions; section B.3 formulates the mathematical model and includes its sensitivity and uncertainty analysis; and section B.4 illustrates and discusses the model validation results.

B.2 Materials and Methods

A demonstration scale biorefinery is capable of transforming approximately $4\,\mathrm{t\,h^{-1}}$ of biomass into three products, i.e. $576\,\mathrm{kg\,h^{-1}}$ 2G bioethanol, $1484\,\mathrm{kg\,h^{-1}}$ C5 molasses and $1740\,\mathrm{kg\,h^{-1}}$ lignin pallets [3]. The biomass can be any lignocellulosic material, preferably agricultural wastes, e.g. straw, bagasses or corn stover, or forest wastes, e.g. saw dust, pulp waste or thinned wood [108]. There are various biomass pretreatment methods, e.g. autohydrolysis, steam explosion, acid hydrolysis etc., among which steam explosion is seen as a cheap and environmentally friendly pretreatment [44]. In this investigation, wheat straw was pretreated with steam in a large pressurized thermal reactor described in [71].

Figure B.1 illustrates the biomass conversion route with emphasis on the enzymatic hydrolysis process. The inflow of pretreated fibers is denoted as F_{FF} and has an initial chemical composition C_0, which is measured with online NIR equipment that performs a measurement every 5 min. A strong base, i.e. NaOH, is pumped into the tank near the inlet for pH adjustments. The base inflow F_B and its concentration C_B are considered known. A certain amount of enzymes proportional to cellulose content from the pretreated fibers is added by a pump from a storage tank. The flow of enzymes F_E and its concentration C_E are considered known. Water is also added with a known flow rate F_W. The reactor is mass controlled using a mass measurement M_{FM} and by manipulating the outflow of fiber mash F_{FM}. The outflow

concentration C is determined by grabbing a sample from the outflow every 6 h and analyzing it in the laboratory with HPLC equipment. The fiber mash pH is measured online with a sampling period of 10 s and all inflows are measured with a sampling period of 2 s.

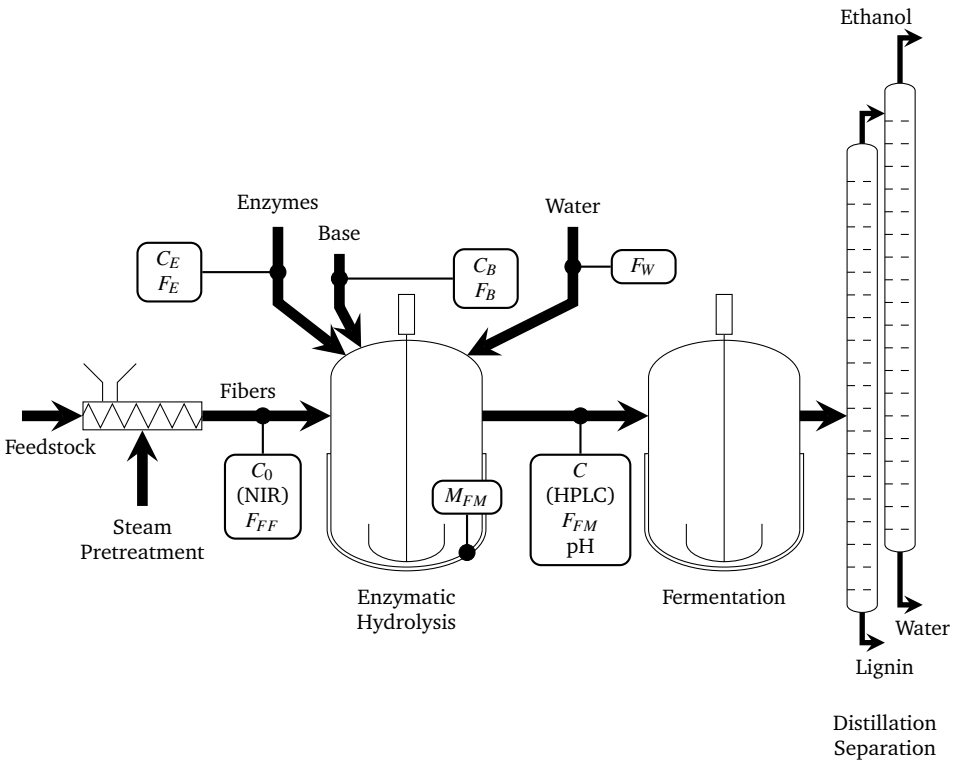


Figure B.1: Biorefinery setup with focus on the enzymatic hydrolysis reactor. All flows are measured, i.e. pretreated fibers F_{FF}, base F_B, water F_W, enzymes F_E and fiber mash F_{FM}. Inflow concentration C_0 is measured by online NIR equipment while C_B and C_E are considered known. There are 2 additional measurements, i.e. a pH sensor on the outflow and a mass indicator M_{FM} for the reactor holdup. The outstream concentration C is determined by analyzing fiber mash samples in the laboratory with an HPLC device.

B.3 Model Development

For this investigation, the mechanistic approach suggested by Kadam et al. [60] is preferred as a starting base. This semi-mechanistic model seems suitable for the

purpose of process operation studies as opposed to the more complex and detailed depolymerization type models [8].

B.3.1 Total Mass Balance

All flows are measured and a total mass balance can be easily constructed:

$$\frac{dM_{FM}}{dt} = F_W + F_{FF} + F_E + F_B - F_{FM} \tag{B.1}$$

B.3.2 Transport Modeling

The pretreatment process creates a slurry with 25 % to 30 % of solid particles, i.e. cellulose, xylan and lignin [55]. The viscosity of this slurry is relatively high in the first phase of the hydrolysis leading to a plug flow transport phenomena in the liquefaction reactor. The container is assumed to have a much larger length than its height, the slurry traversing the tank horizontally from left to right. The mixture is assumed to have uniform properties along the height of the reactor due to vertical mixers.

The species concentration change with respect to time is a combination of convection and diffusion effects plus production or consumption terms. The process normally runs in a continuous manner and slurry transportation is dominated by convection effects. However, sometimes it is necessary to stop the outflow of the tank due to mechanical faults, for example. In such a case, the hydrolysis switches to batch mode and diffusion effects become prominent as cellulosic conversion continues. Convection and diffusion effects are well captured by the Convection-Diffusion-Reaction mass conservation equation [70]:

$$\frac{\partial C}{\partial t} = -\nabla(\vec{v}C) + \nabla(D\nabla C) + R \tag{B.2}$$

where C is the species concentration, $\vec{v}$ is the speed vector, D is the diffusion coefficient and R is the production or consumption rate. The term containing $\vec{v}$ indicates the concentration change due to advection while the term containing D indicates the concentration rate due to diffusion effects. The advection term appears negative because the axis orientation is positively aligned with $\vec{v}$ in the continuity equation. The gradient operator ∇ from Equation (B.2) contains the derivatives only along the x axis due to the vertical uniformity assumption. In this case, the diffusion part of Equation (B.2) is found according to Fick's second law [70]:

$$\nabla(D\nabla C) = \frac{\partial}{\partial x}\left(D\frac{\partial C}{\partial x}\right) \tag{B.3}$$

Substituting Equation (B.3) in (B.2) yields:

$$\frac{\partial C}{\partial t} = -v_x \frac{\partial C}{\partial x} + \frac{\partial}{\partial x}\left(D \frac{\partial C}{\partial x}\right) + R \tag{B.4}$$

Equation (B.4) is solved following a finite element method specific to computational fluid dynamics. Briefly, the reactor is divided into n cells and the space partial derivatives from Equation (B.4) are approximated using a finite volume method as described in [72].

The horizontal speed v_x is considered constant along the length of the reactor and is computed as follows:

$$v_x = \frac{L_R}{t_R} \tag{B.5}$$

where L_R is the length of the liquefaction reactor and t_R is the retention time. The retention time is the ratio between total mass and outflow:

$$t_R = \frac{M_{FM}}{F_{FM}} \tag{B.6}$$

where M_{FM} is the total mass of fiber mash inside the tank and F_{FM} is the outflow.

Diffusion occurs only in the liquid fraction of the slurry and affects only soluble particles, not solids. It is expected to have a varying diffusion along the x axis because the slurry viscosity is changing (expected to decrease) as the liquefaction progresses while liquid viscosity is expected to increase as sugars are formed and dissolved. The Stokes-Einstein equation describes the diffusion coefficient D of solubles as a function of temperature T and liquid viscosity μ_l:

$$D = \frac{kT}{6\pi r \mu_l} \tag{B.7}$$

where k is the Boltzmann constant and r is the molecular radius of the dissolved particle. The radius r of the molecule can be approximated with the following relation:

$$M = N\rho V = N\rho\left(\frac{4}{3}\pi r^3\right) \Rightarrow r = \left(\frac{3M}{4\pi N\rho}\right)^{\frac{1}{3}} \tag{B.8}$$

where M is the molecular mass of soluble component, N is the Avogadro's number, ρ is the solute density and V is its molecular volume. The liquid viscosity μ_l remains to be computed in order to solve Equation (B.7). An empirical relative viscosity equation for high concentration slurries was derived by Thomas [109]:

$$\mu_r = a_0 + a_1\Phi + a_2\Phi^2 \tag{B.9}$$

where Φ is the volume fraction of solid particles and a_1 and a_2 are calibration coefficients estimated as follows by Thomas [109]:

$$a_0 = 1 \qquad a_1 = 2.5 \qquad a_2 = 10.05 \tag{B.10}$$

The relative viscosity μ_r and the liquid viscosity μ_l can then be used to compute the slurry viscosity μ_s:

$$\mu_s = \mu_r \mu_l \tag{B.11}$$

As glucose is produced and dissolved in the liquid part of the slurry, the liquid viscosity changes following a simple linear relation as given by Converti et al. [110]:

$$\mu_l = \mu_W + C_G k_{G\mu} \tag{B.12}$$

where μ_W is the viscosity of the pure liquid solvent, C_G is the glucose concentration and $k_{G\mu}$ is a correlation parameter for glucose. $k_{G\mu}$ and μ_W follow an empirical exponential law also known as the Guzman-Andrade equation [111]:

$$\mu_W = A_W e^{\frac{B_W}{T}} \qquad k_{G\mu} = A_G e^{\frac{B_G}{T}} \tag{B.13}$$

where A_W and B_W are fitting parameters for pure liquid, T is the reactor temperature, and A_G and B_G are parameters for sugar correlation. μ_W is approximately water viscosity and numerical values for A_W, B_W, A_G and B_G are given by Converti et al. [110]:

$$\begin{aligned} A_W &= 2.41 \times 10^{-3}\,\mathrm{g\,m^{-1}\,s^{-1}} \\ B_W &= 1774.9\,\mathrm{K} \\ A_G &= 8.65 \times 10^{-10}\,\mathrm{m^2\,s^{-1}} \\ B_G &= 2502\,\mathrm{K} \end{aligned} \tag{B.14}$$

B.3.3 pH Modeling

The pH model is an extension of the model presented in an earlier publication [12] and comprises a set of 8 weak acids and strong base hydration reactions: self-ionization of water (Equation (B.15c)), acetic acid dissociation (Equation (B.15a)), carbonic acid formation and dissociation (Equations (B.15d) and (B.15e)), succinic acid dissociation (Equations (B.15f) and (B.15g)) and lactic acid dissociation (Equation (B.15h)). The hydration equations are modeled using the approach described by McAvoy et al. [112]. The complete set of equilibrium dissociation equations is

shown next:

$$C_2H_4O_2 \rightleftharpoons C_2H_3O_2^- + H^+ \tag{B.15a}$$
$$NaOH \longrightarrow Na^+ + OH^- \tag{B.15b}$$
$$H^+ + OH^- \rightleftharpoons H_2O \tag{B.15c}$$
$$CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+ \tag{B.15d}$$
$$HCO_3^- \rightleftharpoons CO_3^{2-} + H^+ \tag{B.15e}$$
$$C_4H_6O_4 \rightleftharpoons C_4H_5O_4^- + H^+ \tag{B.15f}$$
$$C_4H_5O_4^- \rightleftharpoons C_4H_4O_4^{2-} + H^+ \tag{B.15g}$$
$$C_3H_6O_3 \rightleftharpoons C_3H_5O_3^- + H^+ \tag{B.15h}$$

Due to the fact that the reactor tank is exposed to atmospheric pressure, it is expected to have a CO_2 stripping process from liquid to gas phase that leads to changes in the carbonic acid and bicarbonate buffers. The CO_2 stripping process or CTR is modeled as in [113] with typical saturation kinetics:

$$CTR = K_{La_{CO_2}} \left({CO_2}^* - CO_{2[aq]} \right) \tag{B.16}$$

where $K_{La_{CO_2}}$ is the mass transfer coefficient, ${CO_2}^*$ is the atmospheric CO_2 concentration and $CO_{2[aq]}$ is the dissolved CO_2 concentration in the reactor.

The dissociation constants for each species at 50 °C, which is a typical optimal reactor temperature for an enzymatic process, are given in Table B.1.

The pH system of equations is solved in each cell of the grid following the procedure from [12]. Thus, the concentration of H^+ is found, which is necessary to compute the pH level:

$$pH = -\log_{10}[H^+] \tag{B.17}$$

Table B.1: Chemical compounds relevant to pH calculation with their dissociation or hydration constants at 50 °C.

Formula	Description	Value	Reference	Variable
$C_2H_4O_2$	Acetic acid	$1.63 \cdot 10^{-5}$	[114]	K_A
H_2O	Water	$5.39 \cdot 10^{-14}$	[114]	K_W
H_2CO_3	Carbonic acid	$5.14 \cdot 10^{-7}$	[114]	K_{C_1}
HCO_3^-	Bicarbonate	$6.69 \cdot 10^{-11}$	[114]	K_{C_2}
$C_4H_6O_4$	Succinic acid	$6.51 \cdot 10^{-5}$	[114]	K_{S_1}
$C_4H_5O_4^-$	Succinate ion	$2.08 \cdot 10^{-6}$	[114]	K_{S_2}
$C_3H_6O_3$	Lactic acid	$1.27 \cdot 10^{-4}$	[114]	K_L

B.3.4 Modeling of Cellulose and Xylan Conversion

The kinetic model is an extension of the work performed by Kadam et al. [60]. Here, the Kadam model is extended taking into consideration enzymes that are capable of decomposing both celullose and xylan, inhibition by furfural of the enzymatic activity, and pH dependency of the enzymatic activity. During xylan hydrolysis, acetic acid is released contributing to lowering the pH level in the mixture. Tracking the acetic acid formation is crucial for building the pH profile of the reactor.

The overall conceptual hydrolysis mechanism is shown in Figure B.2. Cellulose is decomposed into cellobiose and glucose with reaction rates r_1 and r_2, respectively. Cellobiose is further decomposed into glucose with rate r_3. Xylan is hydrolyzed into xylose with an r_4 rate and acetic acid is released as a by-product with an r_5 rate. Cellobiose and glucose formation are inhibited by the sugars themselves or by furfural, which was formed in the pretreatment process.

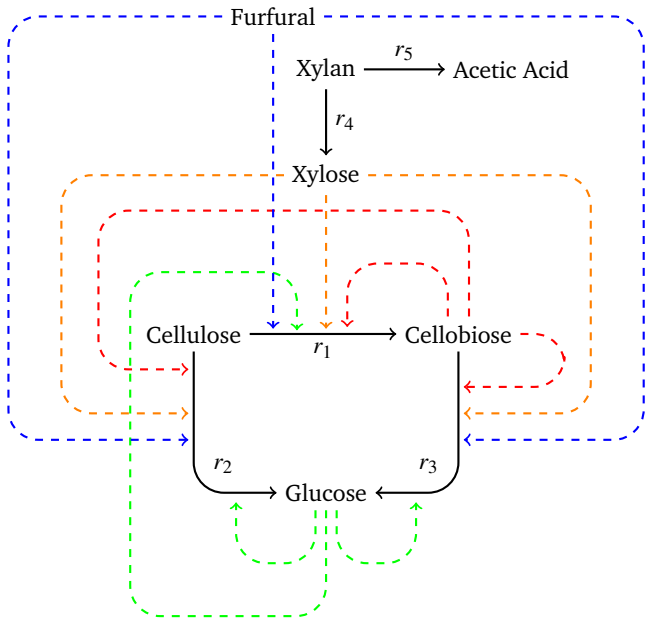


Figure B.2: Conceptual hydrolysis mechanism extended from Kadam et al. [60]. Cellulose is converted by CBH+EG and G into cellobiose and glucose with reaction rates r_1 and r_2, respectively. Cellobiose is converted into glucose with rate r_3. Hemicellulose hydrolysis is illustrated by xylan conversion into xylose with rate r_4, which also releases acetic acid with rate r_5. Dashed lines show inhibition by sugars and furfural (extended from Kadam et al. [60]).

The following chemical species are tracked in the conversion model:

$$C_K = \begin{bmatrix} C_{C_S} & C_{X_S} & C_{L_S} & C_C & C_G & C_X & C_F & C_E \end{bmatrix}^\top \tag{B.18}$$

where C_{C_S}, C_{X_S} and C_{L_S} are solid components from the slurry, i.e. cellulose, xylan and lignin concentrations, and C_C, C_G, C_X, C_F and C_E constitute the liquid part of the slurry, i.e. cellobiose, glucose, xylose, furfural and enzymes, respectively. Equation (B.4) is evaluated for each species from Equation (B.18). The boundary conditions of Equation (B.4) are constructed from the initial chemical composition of pretreated fibers, or C_{K_0}, which is measured in reality with NIR equipment.

The enzymes cocktail is a complex mixture containing many protein types that could be divided into cellulase, hemicellulase and other type of proteins. Cellulose is hydrolyzed by the cellulase group, which is subdivided into exo-β-1.4-cellobiohydrolase and endo-β-1.4-glucanase (CBH+EG), notated as C_{E_1}, and β-glucosidase (G) enzymes, i.e. C_{E_2}. Xylan is hydrolyzed by the hemicellulase (X) group, i.e. C_{E_3}. There are other types of enzymes in the mixture in negligible amounts and their concentration is notated as C_{E_4}. The enzymatic cocktail can then be parametrized in α_i with i from 1 to 4 representing the fraction of each enzyme type. The total enzyme concentration can then be expressed as the sum of each enzyme concentration:

$$C_E = C_{E_1} + C_{E_2} + C_{E_3} + C_{E_4} \tag{B.19}$$

There is an equilibrium between bounded and free enzymes, which is simplified to the following equation:

$$\text{Enzyme}_\text{F} + \text{Solids} \longleftrightarrow \text{Enzyme}-\text{Solids}_\text{B} \tag{B.20}$$

where Enzyme_F are free enzymes and $\text{Enzyme-Solids}_\text{B}$ are bounded enzymes to solids. Therefore, enzymes can be in one of the two states: bounded to solids (B) or free (F). The total concentration of CBH+EG, G or X enzymes consists of bounded and free enzymes:

$$C_{E_i} = C_{E_{i_B}} + C_{E_{i_F}} \tag{B.21}$$

where $i = 1, 2, 3$. The effects of the other type of enzymes, i.e. $i = 4$, is neglected as they do not participate actively to cellulose and xylan decomposition.

The equilibrium between adsorbed and free enzymes follows a Langmuir isotherm expression. In the case of type i enzymes, where i can be CBH+EG, G or X, the Langmuir isotherm relation expresses the ratio between concentration of adsorbed enzymes and solids:

$$\frac{C_{E_{i_B}}}{C_S} = E_{M_i} \frac{K_{A_i} C_{E_{i_F}}}{1 + K_{A_i} C_{E_{i_F}}} \tag{B.22}$$

where $C_{E_{i_B}}$ represents the bounded enzymes, C_S is the total concentration of solids, E_{M_i} is the maximum adsorbed enzymes, K_{A_i} is the adsorption constant and $C_{E_{i_F}}$ is the free enzymes concentration. Cellulose is a percentage of the total solids from the mixture. The concentration of adsorbed enzymes to cellulose $C^C_{E_{i_B}}$ can then be calculated as:

$$C^C_{E_{i_B}} = C_{E_{i_B}} \frac{C_{C_S}}{C_S} = E_{M_i} \frac{K_{A_i} C_{E_{i_F}}}{1 + K_{A_i} C_{E_{i_F}}} C_{C_S} \tag{B.23}$$

where C_{C_S} is the substrate or cellulose concentration for $i = 1, 2$. Similar computations are performed in the case of hemicellulases, or $i = 3$, but the substrate is the xylan concentration this time, or C_{X_S}:

$$C^X_{E_{3_B}} = E_{M_3} \frac{K_{A_3} C_{E_{3_F}}}{1 + K_{A_3} C_{E_{3_F}}} C_{X_S} \tag{B.24}$$

where $C^X_{E_{3_B}}$ represents the concentration of enzymes bound to xylan.

The cellulose to cellobiose reaction rate r_1 with competitive glucose, cellobiose, xylose and furfural inhibition is developed starting from the expression given in Kadam et al. [60] with the extension of pH dependency and furfural inhibition:

$$r_1 = \frac{K_1 \eta_1(T, pH) C^C_{E_{1_B}} C_{C_S}}{1 + \frac{C_C}{I_{C_1}} + \frac{C_X}{I_{X_1}} + \frac{C_G}{I_{G_1}} + \frac{C_F}{I_{F_1}}} \tag{B.25}$$

where K_1 is the reaction constant, $\eta_1(T, pH)$ is the temperature and pH dependency term, I_{C_1} is the cellobiose inhibition term, I_{X_1} is the xylose inhibition term, I_{G_1} is the glucose inhibition term, and I_{F_1} is the furfural inhibition term. Only CBH+EG participate in this reaction.

The cellulose to glucose reaction rate r_2 with competitive glucose, cellobiose, xylose and furfural inhibition is extended similarly:

$$r_2 = \frac{K_2 \eta_2(T, pH) \left(C^C_{E_{1_B}} + C^C_{E_{2_B}} \right) C_{C_S}}{1 + \frac{C_C}{I_{C_2}} + \frac{C_X}{I_{X_2}} + \frac{C_G}{I_{G_2}} + \frac{C_F}{I_{F_2}}} \tag{B.26}$$

where K_2 is the reaction constant, $\eta_2(T, pH)$ is the temperature and pH dependency term, I_{C_2} is the cellobiose inhibition term, I_{X_2} is the xylose inhibition term, I_{G_2} is the glucose inhibition term, and I_{F_2} is the furfural inhibition term. Both CBH+EG and G enzymes participate in this reaction.

The cellobiose to glucose reaction rate r_3 with competitive glucose, cellobiose, xylose and furfural inhibition is extended similarly starting form [60]:

$$r_3 = \frac{K_3 \eta_3(T, pH) C_{E_{2_F}} C_C}{I_3 \left(1 + \frac{C_X}{I_{X_3}} + \frac{C_G}{I_{G_3}} + \frac{C_F}{I_{F_3}} \right) + C_C} \tag{B.27}$$

where K_3 is the reaction constant, I_3 is an overall inhibition term and I_{X_3}, I_{G_3} and I_{F_3} represent inhibition from xylose, glucose and furfural, respectively.

Xylan decomposition is modeled similarly to reaction rate r_1 but with xylan substrate:

$$r_4 = \frac{K_4 \eta_4(T,pH) C^X_{E_{3_B}} C_{X_S}}{1 + \frac{C_C}{I_{C_4}} + \frac{C_X}{I_{X_4}} + \frac{C_G}{I_{G_4}} + \frac{C_F}{I_{F_4}}} \tag{B.28}$$

where r_4 is the xylan to xylose reaction rate, K_4 is the reaction constant, $\eta_4(T,pH)$ represents the temperature and pH dependency, $C^X_{E_{3_B}}$ is the concentration of bounded hemicellulases to xylan, and C_{X_S} is the xylan substrate. I_{C_4}, I_{X_4}, I_{G_4} and I_{F_4} are inhibition terms for cellobiose, xylose, glucose and furfural, respectively.

The temperature and pH dependency can be usually retrieved from the enzymes supplier (e.g. Novozymes). If such data are available the following relation is proposed:

$$\eta_i(T,pH) = \eta_T(T)\eta_{pH}(pH) \tag{B.29}$$

where η_T and η_{pH} are efficiency factors with values between 0 to 1 related to temperature and pH, respectively. The efficiency factors can be either table based or following any arbitrary Gaussian curve functions. Two typical curves for pH and temperature dependency can be found in the supplementary material. These dependencies are implemented as numerical tables and linear interpolation between data points is used throughout simulation.

The amount of acetic acid contained in the hemicellulose part depends on biomass type. For example, in the case of hardwood biomass, 0.5 mol of acetic acid is released for every 1 mol of xylose [115]. Using the molaric mass of xylose and acetic acid, it is found that for every 1 g of xylose, 0.2 g of acetic acid is released. This ratio is parametrized with β and the acetic acid production rate r_5 becomes:

$$r_5 = \beta r_4 \tag{B.30}$$

The predictions of cellobiose, glucose and xylose concentrations are regarded as model outputs:

$$y = [C_C \quad C_G \quad C_X] \tag{B.31}$$

The complete model has 31 parameters: 4 that describe the enzyme mixture composition, i.e. α_i, 1 for the acetic acid to xylose ratio, i.e. β, and 26 parameters relevant for the enzymatic hydrolysis kinetics. α_i and β are fixed a priori and put into vector θ_F. The enzymatic kinetics parameters are gathered into vector θ_K.

B.3.5 Model Calibration, Sensitivity and Uncertainty Analysis

The previous analysis of the Kadam model performed by Sin et al. [8] indicated that the model was over parameterized with respect to available measurements. Hence, for a proper model calibration, a sensitivity analysis was performed to find out the most significant parameters to be used as subset for fine tuning of the model fits to the available data. The sensitivity analysis is conducted by calculating a measure called δ_k^{msqr} as in the methodology described in [64]:

$$\delta_k^{msqr} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(s_{nd,ik}\right)} \tag{B.32}$$

where k is the parameter index in the hydrolysis parameter vector θ_K, i is the model output index and $s_{nd,ik}$ is the non dimensional sensitivity defined as:

$$s_{nd,ik} = \frac{\partial y_i}{\partial \theta_k}\frac{\theta_k}{sc_i} \tag{B.33}$$

where $\partial y_i/\partial\theta_k$ represents the output variation with respect to a variation in parameter θ_k and sc_i is a scaling factor with the same physical dimension as the corresponding observation. All parameters are ranked according to δ_k^{msqr} and a subset θ_S is built with the most significant parameters.

A model calibration follows, which adjusts the parameters from the θ_S subset in order to obtain a better fit. For this calibration, only 50 h out of the 170 h of recorded data are used, while the remaining 120 h data are used for validation. It is noted that the calibration of the model parameters is meant to be a fine-tuning around the nominal values of the model parameters estimated from batch assays by Kadam et al. [60]. A comprehensive system identification is not pursued due to the fact that available plant measurements were obtained from closed loop operation and not under proper optimal experimental design.

Uncertainty Analysis

The uncertainty analysis is carried out using the engineering standard Monte Carlo technique, which includes the following four steps [9]: (1) define input uncertainties with their range; (2) sampling of kinetics and feed parameters using the Latin hypercube sampling with correlation control; (3) run Monte Carlo simulations with sampled values; (4) evaluate results. In the last step, in addition to inference statistics, a sensitivity analysis was also performed using linear regression of Monte Carlo outputs, also known as the standardized regression coefficients (SRC). The methodology is detailed in [9].

In step 1, the uncertainty analysis considers two sources of uncertainty, namely kinetic parameters and feed composition (mass). For definition of the kinetic

parameter uncertainty, a uniform distribution with the corresponding lower and upper bounds are used as defined in [9], while the correlation matrix between the parameters was taken from [8]. Bias and standard deviation of measurements due to sensors were considered for the characterization of feed composition uncertainty. The feed measurements are subject to errors either due to a miss calibration of the NIR equipment or because of a wrongly placed sensor, which often indicates offsets in the mass inflow of fibers. The most relevant feed variables are cellulose and xylan content, and fiber mass inflow. The error that affects these signals are statistically characterized using normal distributions with the following parameters:

$$\varepsilon_C \in N(\mu_C, \sigma_C^2) \quad \varepsilon_X \in N(\mu_X, \sigma_X^2) \quad \varepsilon_F \in N(\mu_F, \sigma_F^2) \tag{B.34}$$

where ε_C, ε_X and ε_F are cellulose, xylan and inflow errors, and N is the normal distribution. The mean values μ_C, μ_X and μ_F represent sensor offsets and are normally distributed such that to cover a range of 5 % to 10 % of their nominal operational values, an error range assumed for NIR equipments. The standard deviations of the measurements are considered to follow gamma distributions, which is good practice in measurement error modeling. Hence the uncertainty on the measured feed composition are generated by performing LHS on 2 parameters of the normal distribution: one that characterizes the probability of the mean value, notated as F_μ, and another one for the standard deviation, i.e. F_σ.

The SRC method fits a linear regression model to Monte Carlo simulation outputs [65]:

$$y_{reg} = b_0 + \sum_{i=1}^{N_\theta} b_i \theta_i \tag{B.35}$$

where θ_i is the uncertain parameter vector of length N_θ, b_i are the regression coefficients and y_{reg} is the fitted model output. The regression coefficients b_i are then scaled with respect to the standard deviation in parameters and simulation outputs:

$$\beta_i = \frac{\sigma_{\theta_i}}{\sigma_y} b_i \tag{B.36}$$

where β_i are the standardized regression coefficients, σ_{θ_i} is the standard deviation of parameter uncertainty and σ_y is the simulation output standard deviation.

B.3.6 Simulation Scenarios

The model is evaluated in two different scenarios: the first one concerns a steady-state average performance simulation while in the second one the model is driven by real measurements and its predictions are compared against HPLC data throughout 170 h of operation.

In the first scenario, the reactor is assumed to be filled initially with water and all model inputs, including feed composition, are set to a typical operational point of a demonstration scale biorefinery as in Table B.2. The reactor is capable of processing $4\,\mathrm{t\,h^{-1}}$ of fibers. Water and enzymes inflows are adjusted proportionally to fibers inflow. A mass controller automatically sets the outflow of fiber mash in order to keep a constant holdup. The purpose of this scenario is to observe the steady-state concentrations, pH, viscosity and solids profiles from the reactor.

Table B.2: Dry matter composition of steam pretreated wheat straw at 170 °C: demonstration scale plant values against composition determined by [116].

Component	Demonstration Plant	[116]
Cellulose	45.0 %	50.0 %
Xylan	10.0 %	13.6 %
Lignin	32.0 %	20.8 %
Other	13.0 %	15.6 %

B.4 Results and Discussion

B.4.1 Model Sensitivity and Calibration

Model parameters are initialized to reported values from [60], which are given in Table B.3. There are a few missing parameters referring to xylan hydrolysis and furfural inhibition, which have not been reported earlier in the literature. These parameters are initialized to reasonable values.

The δ^{msqr} sensitivity measure is numerically computed for each output as δ_C^{msqr}, δ_G^{msqr} and δ_X^{msqr}, representing cellobiose, glucose and xylose, respectively, and also as an overall indicator δ^{msqr}. Parameters are ranked with respect to δ^{msqr} and the results are shown in Figure B.3. As expected, the model is over parametrized and only a subset of parameters is relevant for the output dynamics. A threshold for δ^{msqr} is set to 0.025 in this case (there is no absolute threshold value reported [8]) and shown in the bottom plot of Figure B.3. The subset of significant parameters is found as follows:

$$\theta_S = [K_1 \quad K_2 \quad K_4 \quad E_{M_1} \quad K_{A_1} \quad I_{C_1} \quad I_{G_1} \quad I_{X_1} \quad I_{G_2} \quad I_{X_2} \quad I_{G_4}]^\top \tag{B.37}$$

The sensitivity analysis results indicate that the reaction rates that participate in cellobiose, glucose and xylose formation, i.e. K_1, K_2 and K_4, are influential parameters. Inhibition of r_1 and r_2 by glucose and xylose, i.e. I_{G_1}, I_{G_2}, I_{X_1} and I_{X_2}, as well as I_{G_4} or inhibition of r_4 by glucose, and I_{C_1} or inhibition of r_1 by cellobiose, also

constitute important model parameters. The cellullase adsorption parameters E_{M_1} and K_{A_1} are the parameters closest to the threshold indicating a slight significance. The furfural inhibition parameters, i.e. I_{F_1}, I_{F_2}, I_{F_3} and I_{F_4} have low sensitivity on model outputs, which is in accordance to [52] who observed a reduced inhibitory effect of furfural in the enzymatic hydrolysis process. All other model parameters, i.e. the ones that influence the cellobiose to glucose reaction r_3, and cellobiose inhibition on the other reactions, have almost no sensitivity on model outputs.

The parameters in θ_S are calibrated on a reduced set of recorded data in order to obtain a better fit. The calibrated values are shown in the second column of Table B.3. As indicated by Sin et al. [8], there is a high correlation between parameters and a large confidence interval for their numerical values. Therefore, a physical meaning cannot be attached to these parameters. This explains why some calibrated parameters are significantly different than those reported by [60].

Table B.3: Model parameters initialization and calibration.

Parameter	Value	[60]
θ_F	Independently fixed parameters	
α_1	0.5	-
α_2	0.3	-
α_3	0.2	-
α_4	0	-
β	0.2	-
θ_S	Calibrated parameters	
K_2	0.0053	0.002
E_{M_1}	0.015	0.06
I_{X_2}	0.029	0.2
K_1	0.00034	0.0062
K_4	0.0027	-
I_{C_1}	0.0014	0.015
I_{G_4}	2.39	-
I_{G_1}	0.073	0.1
K_{A_1}	0.84	0.4
I_{G_2}	0.34	0.04
I_{X_1}	0.1007	0.1
Parameters with low sensitivity		
E_{M_2}	0.01	0.01
E_{M_3}	0.01	-
K_{A_2}	0.1	0.1

K_{A_3}	0.1	-
I_{F_1}	10	-
I_{C_2}	132	132
I_{F_2}	10	-
I_3	24.3	24.3
I_{X_3}	201	201
I_{G_3}	3.9	3.9
I_{F_3}	10	-
I_{C_4}	24.3	-
I_{X_4}	201	-
I_{F_4}	10	-
K_3	0.07	0.07

B.4.2 Steady State Average Performance Simulation Scenario

The model inputs are set to the constants given in Table B.4, which represent an average performance operating point of the biorefinery. The percentage values from Table B.2 for the demonstration plant are transformed into $\mathrm{g\,kg^{-1}}$ in Table B.4, given the fact that the fibers have a 25 % dry matter. The CO_2 concentration is set to the saturation point, i.e. $0.0011\,\mathrm{g\,kg^{-1}}$ at 50 °C. The simulation runs for 20 h and the stabilized reactor profiles can be observed in Figure B.4 drawn with a solid black line.

Subplot A illustrates the buffer concentrations in $\mathrm{mol\,L^{-1}}$ relevant for pH calculations. The acetic acid concentration, notated as A_T, increases along the reactor due to xylan hydrolysis, as expected. The base concentration B_T, the bicarbonate C_T, the succinic acid S_T and lactic acid L_T are constant throughout the reactor as there is no production or consumption of these species.

Subplot B shows the conversion of cellulose and xylan into cellobiose, glucose and xylose. Cellulose C_{C_S} and xylan C_{X_S} drop along the reactor as the liquefaction process progresses. As sugars are formed, the conversion rate decreases due to the inhibition effects of newly created xylose and glucose, which is in accordance with Equations (B.26) and (B.28). This is why glucose and xylose are formed more rapidly in the first sections of the reactor. Approximately 20 % of the entering cellulose is converted into glucose for the given reactor.

Subplots C, D and *E* illustrate the pH, viscosity and solids profiles. Acetic acid formation leads to a drop in pH level of 0.4 units along the reactor. This pH deviation affects the optimal performance of the enzymes, which also contributes to a slower

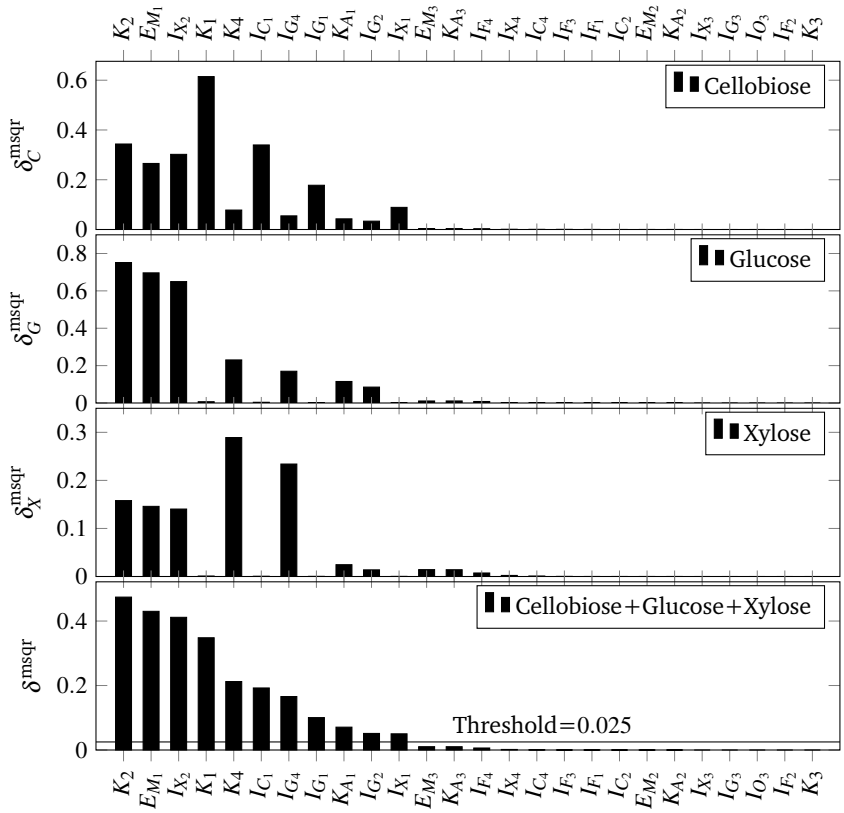


Figure B.3: Sensitivity measure δ^{msqr} of model outputs with respect to kinetics parameters θ_K. The first 3 plots show δ^{msqr} for cellobiose (δ_C^{msqr}), glucose (δ_G^{msqr}) and xylose (δ_X^{msqr}). The bottom plot illustrates the overall sensitivity measure δ^{msqr} and a threshold that delimits the relevant parameters subset θ_S from the rest of the parameters.

conversion. Monitoring the pH profile can tell how optimal a hydrolysis process runs from the point of view of enzymatic activity. The slurry viscosity drops along the reactor by 0.15 units and is a good indicator of how liquid the mixture becomes, i.e. viscous or not. The solids profile shows a reduction of 3 % in solids, mainly due to cellulose and xylan liquefaction. The other solids, e.g. lignin, remain in the mixture unchanged. The reactor retention time in this simulation scenario is 7.8 h, which is enough time to create a transportable slurry that can be pumped to subsequent tanks for continuing the hydrolysis process. A typical hydrolysis process requires 140 h to 160 h [60].

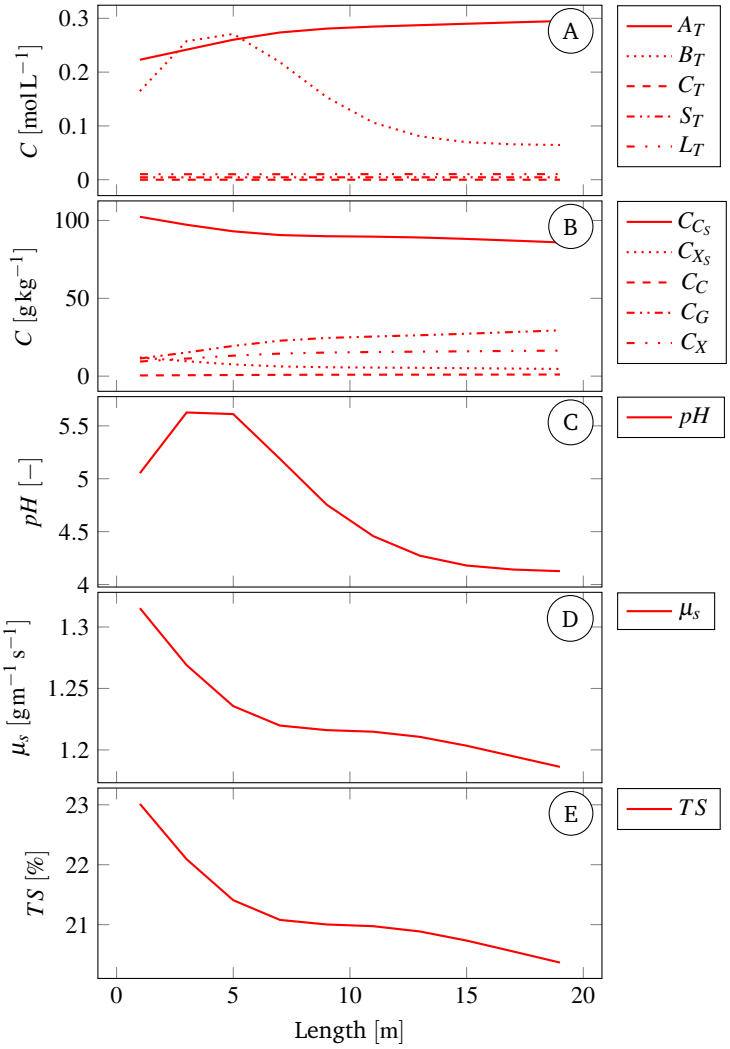


Figure B.4: Reactor profiles from the average performance simulation scenario with constant inputs (drawn with ——) and from time $t = 24$h when the model is driven by real measurements (drawn with ——). Subplot A shows the buffer concentrations relevant for pH calculations, i.e. acetic acid A_T, base B_T, bicarbonate C_T, succinic acid S_T and lactic acid L_T. Subplot B illustrates cellulose C_{C_s} and xylan C_{X_s} conversion into cellobiose C_C, glucose C_G and xylose C_X. Subplots C, D and E display the pH, viscosity and solids profiles along the reactor.

Table B.4: Boundary conditions for a typical demonstration scale process operation used in the average performance simulation scenario.

Input	Description	Value	Unit
F_{FF}	Fiber flow	1.11	$\mathrm{kg\,s^{-1}}$
F_B	Base flow	0.012	$\mathrm{kg\,s^{-1}}$
F_E	Enzymes flow	0.025	$\mathrm{kg\,s^{-1}}$
F_W	Water flow	0.014	$\mathrm{kg\,s^{-1}}$
F_{FM}	Fiber mash outflow	1.16	$\mathrm{kg\,s^{-1}}$
C_E	Enzymes concentration	500.0	$\mathrm{g\,kg^{-1}}$
C_{B_0}	Base concentration	270	$\mathrm{g\,kg^{-1}}$
	Fiber fraction composition		
$C_{C_{S_0}}$	Cellulose concentration	112.5	$\mathrm{g\,kg^{-1}}$
$C_{X_{S_0}}$	Xylan concentration	20	$\mathrm{g\,kg^{-1}}$
$C_{L_{S_0}}$	Lignin concentration	80	$\mathrm{g\,kg^{-1}}$
$C_{A_{T_0}}$	Acetic acid concentration	5.0	$\mathrm{g\,kg^{-1}}$
$C_{C_{T_0}}$	CO_2 concentration	0.0011	$\mathrm{g\,kg^{-1}}$
$C_{S_{T_0}}$	Succinic acid concentration	0.4	$\mathrm{g\,kg^{-1}}$
$C_{L_{T_0}}$	Lactic acid concentration	0.7	$\mathrm{g\,kg^{-1}}$
C_{G_0}	Glucose concentration	0.5	$\mathrm{g\,kg^{-1}}$
C_{C_0}	Cellobiose concentration	0.0	$\mathrm{g\,kg^{-1}}$
C_{X_0}	Xylose concentration	2.5	$\mathrm{g\,kg^{-1}}$
C_{F_0}	Furfural concentration	1.8	$\mathrm{g\,kg^{-1}}$
T	Optimal enzymatic activity temperature	50.0	°C

B.4.3 Dynamic Simulation Scenario with Full Scale Real Measurements and Uncertainty Analysis

Real data from a demonstration scale biorefinery have been recorded throughout 170 h of operation. The model is simulated by using the recorded plant online data including feed composition and flux. A snapshot of the reactor profiles is caught at time $t = 24\,\mathrm{h}$. The profiles are drawn with red lines in Figure B.4 overlapped with the simulation scenario profiles. In *subplot A* it is seen that real data show a higher concentration of acetic acid in the reactor, which causes a lower enzymatic activity. More base starts to be pumped (shown with dotted red line in *subplot A*) in order to adjust the pH level and an overshoot is recorded as the pH level increases to 5.5 units near the inlet (*subplot C*). Still, the overall enzymatic activity is improved and a conversion close to the theoretical profile found in the previous scenario is recorded. Consequently, the slurry viscosity μ_s and the total solids percentage start to drop as

illustrated in *subplots D* and *C*. Conversion still occurs even though the hydrolysis process runs suboptimal from the point of view of the enzymatic activity.

The top plot in Figure B.5 illustrates the outstream pH prediction, which fits relatively well with the sensor data. There are several reasons that can cause estimation differences: the online HPLC analyzer of pretreated fibers is the major source of inaccuracies regarding the relevant buffer concentrations for pH calculations; the flow measurement of fibers also presents drifts and offsets leading to an erroneous calculation of the incoming acids affecting the estimated pH; estimation errors could also occur due to unknown buffers created in the pretreatment process; a last source of errors could be the pH sensor itself, which can have offsets if the cleaning procedure is not followed properly. For example, it is suspected that the pH sensor has an offset error between 90 h to 140 h because the pH measurement drops below 4 units, which is unlikely to happen in reality. Also, the measurements in this time frame are very noisy compared to the non-faulty case and there is a sudden change in the measurement at approximately $t = 140\,\text{h}$, which can happen only if the sensor has been cleaned at that specific time.

The pH estimation is more dynamic than the measurement, presenting several spikes when the reactor holdup is changing but stays most of the time within 4 to 5 pH units with an error below 1 unit, which is considered satisfactory knowing the complexity of the chemical species in the slurry.

The bottom plot from Figure B.5 shows the calculated fiber mash viscosity or μ_S. The values indicate a rather liquid mixture that can be pumped further to subsequent tanks. Typically, the viscosity near the tank inlet is above 1.3 units as indicated in *subplot D* from Figure B.4.

Throughout the operation of the biorefinery, samples were grabbed at the outflow every 6 h. These samples were then analyzed with an HPLC device in the laboratory in order to observe the sugar concentrations. Figure B.6 displays glucose, xylose and cellobiose concentrations from the outstream, i.e. measurements with bullet points and predicted values with solid lines. The gray area indicates the 5th-95th percentile interval obtained after running the Monte Carlo simulations with LHS sampling on kinetics and feed parameters as explained in the methodology. The sugar formation fits fairly well within the 5th-95th percentile. Some differences appear when the reactor holdup changes, which can be due to non trivial mixing effects that are not captured well by the 1-D transport equation.

The SRC coefficients are summarized in Table B.5, where all kinetic and feed parameters are ranked with respect to the β_i coefficients. The ranking shows that most of the model output uncertainty is explained by the uncertainties in the kinetic parameters. The contribution of the feed composition uncertainties to the model

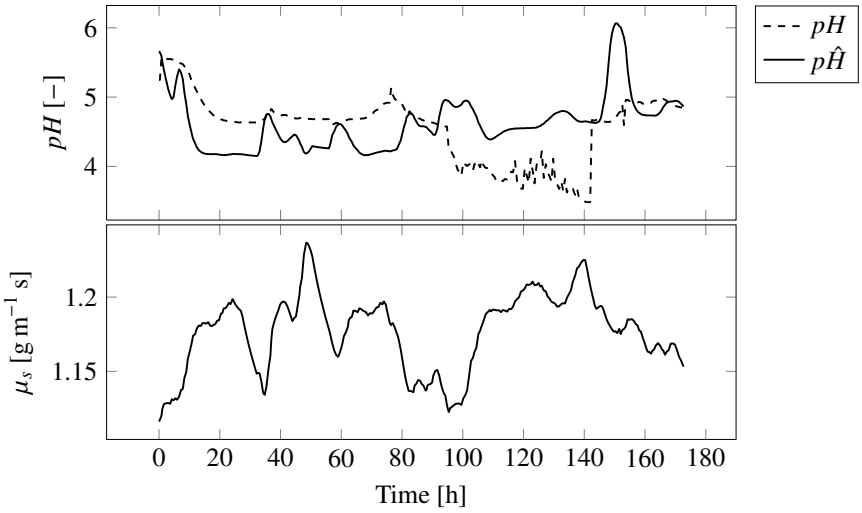


Figure B.5: pH and viscosity calculations for the reactor outstream. The top plot compares the calculated pH, i.e. $p\hat{H}$, against the real measurement pH. The bottom plot shows the calculated slurry viscosity.

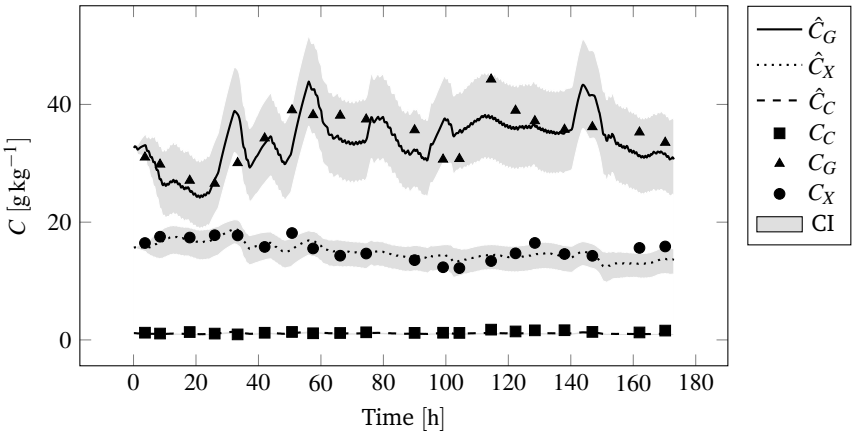


Figure B.6: Cellobiose, glucose and xylose formation during the enzymatic hydrolysis process. $\hat{C}_C$, $\hat{G}_C$ and $\hat{X}_C$ are estimated concentrations of cellobiose, glucose and xylose as predicted by the model while C_C, C_G and C_X are sugar concentrations measured by the HPLC equipment. The gray area shows the 5th-95th percentile interval considering kinetics and feed uncertainties.

output are much smaller. This makes sense because the hydrolysis tank has large volume and acts as a buffer to feed variations while the intrinsic uncertainties of the enzyme kinetics affect the degree of conversion of cellulose, hence the predicted performance of the reactor.

Table B.5: Parameter ranking with respect to the SRC coefficients: β_{i_C}, β_{i_G} and β_{i_X} are the standardized regression coefficients for cellobiose, glucose and xylose, respectively.

Rank	Parameter	β_{i_C}	Parameter	β_{i_G}	Parameter	β_{i_X}
1	K_1	1.00	I_{G_2}	1.00	I_{G_2}	0.99
2	I_{G_2}	1.00	I_{X_1}	0.99	K_2	0.95
3	I_{G_1}	1.00	K_2	0.99	K_4	0.68
4	I_{C_1}	1.00	I_{G_1}	0.72	I_{G_4}	0.56
5	I_{X_2}	0.66	E_{M_1}	0.71	F_μ	0.31
6	I_{X_1}	0.42	K_1	0.60	E_{M_1}	0.21
7	K_2	0.27	K_4	0.42	I_{X_1}	0.17
8	E_{M_1}	0.25	I_{C_1}	0.33	K_1	0.16
9	K_4	0.24	I_{G_4}	0.28	I_{X_2}	0.08
10	F_μ	0.17	K_{A_1}	0.21	I_{C_1}	0.07
11	K_{A_1}	0.15	F_μ	0.18	K_{A_1}	0.06
12	I_{G_4}	0.08	F_σ	0.06	I_{G_1}	0.04
13	F_σ	0.01	I_{X_2}	0.04	F_σ	0.03

B.4.4 Perspectives

The presented model of the hydrolysis reactor is a promising tool for simulation based process studies. One example is monitoring using soft sensors. A soft sensor is an application of the formulated dynamic model that is able to provide information about variables of interest that cannot be directly measured or, for which, sensors are missing. In the hydrolysis reactor case, the pH and viscosity profiles cannot be directly measured but can be constructed by the model. These profiles constitute important monitoring tools for assessing the performance of the enzymatic process.

Another application deals with fault diagnosis. Algorithms can be developed that investigate the differences between the model predictions and the real measurements in order to detect and isolate any faults that drifts the process from running normally. For example, it could be automatically found when the pH sensor would need to be cleaned.

Process control could also be performed with the current model. Advanced observer based control strategies can be derived for controlling the pH level, the

glucose formation or enzyme dosage. Other possibilities will be investigated in the future.

While many promising applications are possible, one area of improvement is in the calculation of viscosity and its comparison to real-measurements which is subject to ongoing work. Additionally, given how uncertainty in model parameters affects the prediction quality, robust techniques can be used for both control and optimization purposes. These are subject to ongoing work.

B.5 Conclusions

This investigation formulated and analyzed an enzymatic hydrolysis dynamic model for demonstration scale processes. The model consists of a transport module, coupled viscosity estimation and dynamic pH prediction, and pH dependent enzymatic kinetics. The uncertainty analysis performed on the model predictions indicated an acceptable variance that matches well with the measured variability of glucose, xylose and cellobiose of long term plant data. This indicates the quality and reliability of the model as a valuable tool for monitoring, diagnosis and control design.

Acknowledgments

The close collaboration and very helpful suggestions and detailed comments received from Dr. Jakob M. Jensen and Eng. Michael Elleskov on this research are gratefully acknowledged.

Supplementary Material

Finite Volume Method for CDR Equation

Equation (4) is discretized in space and appropriate boundary conditions are set, i.e. Neumann type on the western border and Dirichlet type at east [72]. To perform the space discretization, the reactor is sliced along the x axis as in Figure B.7, creating cells of width δx.

Each unit P has two neighbors, i.e. W to the west and E to the east. The border between the slices is denoted as w and e, respectively. Following a finite volume method, Equation (4) is integrated in the volume unit:

$$\int_{\Delta V} \frac{\partial C}{\partial t} dV = -v_x \int_{\Delta V} \frac{\partial C}{\partial x} dV + \int_{\Delta V} \frac{\partial}{\partial x} \left(D \frac{\partial C}{\partial x} \right) dV + \int_{\Delta V} R dV \tag{B.38}$$

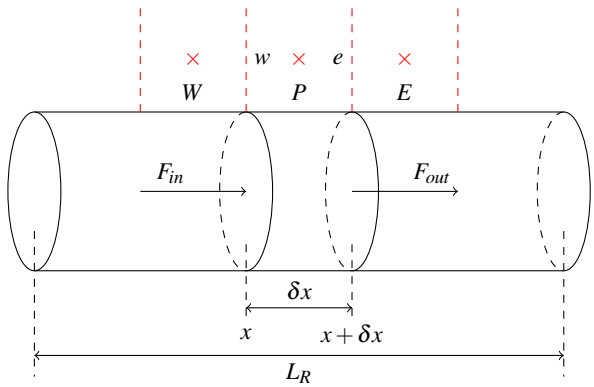


Figure B.7: Discretization of the reactor in cells along the x axis. F_{in} is the mass inflow, F_{out} is the cell outflow and δx is the cell width. P denotes the current cell, W is its western neighbor while E denotes the eastern cell. w and e indicate the western and eastern borders, respectively.

Enzymatic Activity Dependency

Figure B.8 shows the dependency of enzymatic activity on pH and temperature as it appears in the enzyme data sheet [58].

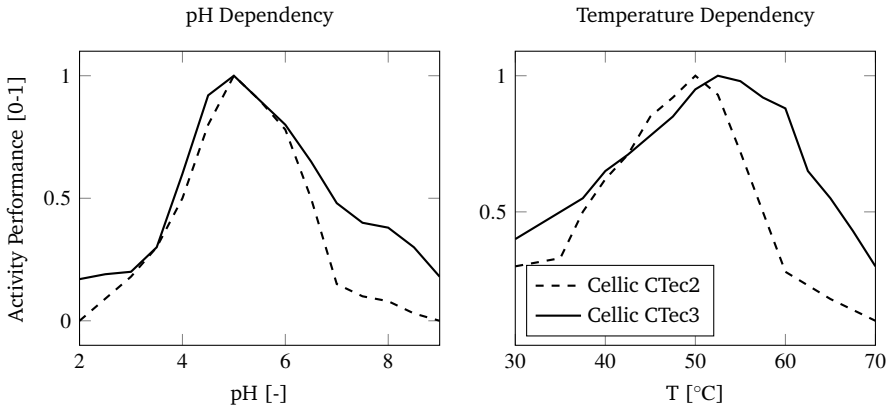


Figure B.8: Enzymatic activity dependency on pH and T [58]. These dependencies resemble Gaussian curves and are implemented as numerical tables with linear interpolation for simulation.

Parameter Uncertainty Definition for Monte Carlo Simulations

Table B.6 shows the uncertainty in kinetics parameters that was considered in the investigation for generating the Latin hypercube samples.

Most uncertainty definitions are adapted from [9] while the uncertainty definition for the kinetic parameters related to hemicellulose hydrolysis and furfural inhibition are defined assuming a class 2 uncertainty, i.e. 25 % uncertainty around the nominal value with uniform distribution.

The correlation matrix between kinetic parameters is adapted from [8]. This is used within the Latin hypercube sampling to introduce correlation control.

Latin Hypercube Sampling on Feed Parameters

Figure B.9 shows 5 feed scenarios after performing the Latin hypercube sampling on the feed parameters, i.e. bias and standard variation.

Table B.6: Expert review of uncertainty for kinetics parameters adapted from [9].

Parameter	Default Value	Lower Bound	Upper Bound
K_1	0.00034	0.000255	0.000425
K_2	0.0053	0.00398	0.00663
K_4	0.0027	0.00203	0.00338
E_{M_1}	0.015	0.0113	0.0188
K_{A_1}	0.84	0.63	1.05
I_{C_1}	0.0014	0.00105	0.00175
I_{G_1}	0.073	0.0548	0.0913
I_{G_2}	0.34	0.255	0.425
I_{G_4}	2.39	1.79	2.99
I_{X_1}	0.101	0.0755	0.126
I_{X_2}	0.029	0.0218	0.0363

Table B.7: Correlation matrix for the Latin hypercube sampling from the kinetic parameter uncertainty space.

	K_2	E_{M_1}	I_{X_2}	K_1	K_4	I_{C_1}	I_{G_4}	I_{G_1}	K_{A_1}	I_{G_2}	I_{X_1}
K_2	1	0	−1.00	0.29	0	−0.29	0	−0.29	0	−1	−0.29
E_{M_1}	0	1	0	0	0	0	0	0	0	0	0
I_{X_2}	−1.00	0	1	−0.29	0	0.29	0	0.29	0	1.00	0.29
K_1	0.29	0	−0.29	1	0	−1.00	0	−1	0	−0.29	−1.00
K_4	0	0	0	0	1	0	0	0	0	0	0
I_{C_1}	−0.29	0	0.29	−1.00	0	1	0	1.00	0	0.29	1.00
I_{G_4}	0	0	0	0	0	0	1	0	0	0	0
I_{G_1}	−0.29	0	0.29	−1	0	1.00	0	1	0	0.29	1.00
K_{A_1}	0	0	0	0	0	0	0	0	1	0	0
I_{G_2}	−1	0	1.00	−0.29	0	0.29	0	0.29	0	1	0.29
I_{X_1}	−0.29	0	0.29	−1.00	0	1.00	0	1.00	0	0.29	1

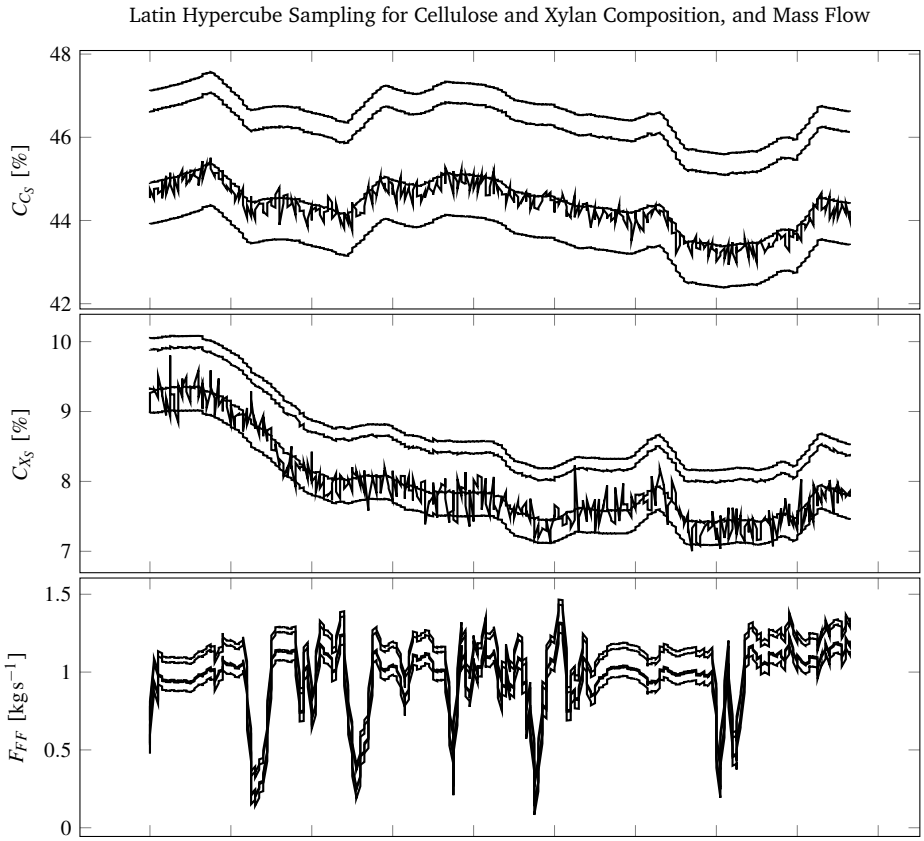


Figure B.9: Model inputs generated after Latin hypercube sampling for feed parameters, i.e. F_μ and F_σ. The bias and the variance vary depending on the Latin hypercube samples. In this figure, only 5 samples are plotted for the purpose of visualization while 250 samples were generated for the Monte Carlo simulations presented in the manuscript. C_{C_S} is the cellulose composition, C_{X_S} is the xylan content and F_{FF} is the mass flow of fibers.

Paper C

Model-based Plantwide Optimization of a Large Scale Lignocellulosic Bioethanol Plant

Remus Mihail Prunescu[1], Mogens Blanke[1], Jon Geest Jakobsen[2], Gürkan Sin[*3]

[1]Department of Electrical Engineering, Automation and Control Group, Technical University of Denmark, Elektrovej Building 326, 2800, Kgs. Lyngby, Denmark

[2]Department of Process Control and Optimization, DONG Energy Thermal Power A/S, Nesa Allé 1, 2820, Gentofte, Denmark

[3]CAPEC-PROCESS, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads Buildings 227 and 229, 2800, Kgs. Lyngby, Denmark

Abstract:
Second generation biorefineries transform lignocellulosic biomass into chemicals with higher added value following a conversion mechanism that consists of: pretreatment, enzymatic hydrolysis, fermentation, and purification. The objective of this study is to identify the optimal operational point with respect to maximum economic profit of the large scale biorefinery plant using a systematic model-based plantwide optimization methodology. The key process parameters or decision variables are identified as: pretreatment temperature, enzyme dosage in enzymatic hydrolysis, and yeast loading per batch in fermentation. The plant is treated in an integrated manner taking into account the interactions and trade-offs between the conversion steps. A sensitivity and uncertainty analysis follows for the optimal solution considering both model and feed parameters. It is found that an optimization supervisory layer is superior to a traditional refinery operation, and also reduces the uncertainty on the profit curve.

[*]Principal corresponding author. Tel.: +45 45252806; E-mail: gsi@kt.dtu.dk

C.1 Introduction

Second generation lignocellulosic biorefineries reached commercial reality in 2012 [3], and consequently many large scale plants are in operation nowadays: e.g. Beta Renewables, Abengoa Bioenergy, GranBio, POET-DSM [2]. Most biorefineries produce bioethanol but the drop in oil price reduced the demand on the biofuel. However, plant upgrades for chemicals with higher-added values are recommended making biorefineries still competitive in an oil dependent environment [26].

This study deals with optimizing the operation of a large scale second generation biorefinery with a well established conversion route for bioethanol production using modeling and simulation as enabling technology. The objective of the optimization problem is to maximize the plant economical profit, considering prices for the most important inputs and outputs: biomass, steam, enzymes, yeast, and ethanol among others.

The conversion route from lignocellulosic material to products with higher added value consists of: pretreatment, enzymatic hydrolysis, fermentation, and purification [3, 4]. Lignocellulosic biomass contains cellulose, hemicellulose (xylan and arabinan), lignin, ash, and other residues [37]. The scope of the pretreatment process is to open the biomatrix, relocate lignin and partially hydrolyze the hemicellulose such that cellulose would become more accessible to the downstream process of enzymatic hydrolysis [40]. During pretreatment, inhibitors such as organic acids, furfural, and 5-Hydroxymethylfurfural (5-HMF) are also created due to sugar degradation. Organic acids change the pH of medium, but can be automatically neutralized by a pH controller for ensuring optimal enzymatic conditions [11]. Furfural, 5-HMF, and acetate are fermentation inhibitors [52], while the remaining hemicellulose fraction leads to xylooligomers and xylose formation in the enzymatic hydrolysis process, which strongly inhibit the enzymatic activity [51].

There are trade-offs between the conversion steps. Too little biomass pretreatment would reduce the exposed cellulose to enzymes, and also increases the amount of hemicellulose for enzymatic hydrolysis, which would eventually decrease the glucose yield due to xylose and xylooligomers inhibition. On the other hand, too much biomass pretreatment would increase the amount of fermentation inhibitors leading to a lower ethanol yield.

Most existing studies focus on operational optimization conducting small scale experiments in the laboratory for finding the best pretreatment conditions such that ethanol yield is maximized [50, 116, 117, 118]. The focus is on one unit at a time (pretreatment versus enzymatic hydrolysis versus fermentation) but rarely consider the steps simultaneously in an integrated manner. Such methods are time

consuming, suboptimal from an economic point of view as they do not take into account process economics, and not suitable for large scale facilities. Also, in these studies, the enzymatic hydrolysis and fermentation processes are usually conducted following a fixed recipe, i.e. no correction action or feedback is taken to counteract the effects of inhibitors. For example, one could increase the enzyme dosage when xylooligomers and xylose inhibit glucose production, or adjust the yeast seed in fermentation to compensate for inhibitors.

Therefore there is a need for systematic methods and tools to facilitate the further process optimization and operation of second generation bioethanol plants, which is the objective of this study. Moreover, the optimal solution proposed is also analyzed from a sensitivity and uncertainty point of view with respect to feedstock composition and kinetic parameters. A Monte Carlo technique with Latin Hypercube Sampling and correlation control is used for the uncertainty analysis following the methodology from [9, 66].

This paper is structured as follows: the methods section revises the methodology for building the optimization layer for plantwide operation, along with the theoretical part of the sensitivity and uncertainty analysis as well as the development of necessary plant models. The results and discussion follow where the profit curve, costs, and optimal solutions are presented along with their uncertainty bounds. The paper concludes with a summary of all important findings.

C.2 Methods

C.2.1 Second Generation Bioethanol Plant

Figure C.1 illustrates the second generation biorefinery concept for bioethanol production. The pretreatment process consists of a continuous thermal reactor and a separation press, which were modeled and analyzed in [5, 71]. The thermal reactor is equipped with a temperature controller for adjusting the reaction temperature T_{tr} [12]. When hemicellulose is hydrolyzed, it produces xylose and arabinose (C5 sugars). After separation, the liquid part containing the C5 sugars is directly pumped into fermentation reactors, bypassing the enzymatic hydrolysis reactors. Cellulose can also be degraded in the pretreatment process, but the produced glucose (C6 sugar) is not lost as it is added to fermentation along with the C5 sugars from the liquid fraction.

The enzymatic hydrolysis process was thoroughly described and analyzed in [6]. It runs at a high dry matter content in a continuous mode and consists of a series of hydrolysis tanks. The first reactor is described in [54] followed by conventional continuous stirred tank reactors in order to meet the necessary hydrolysis retention

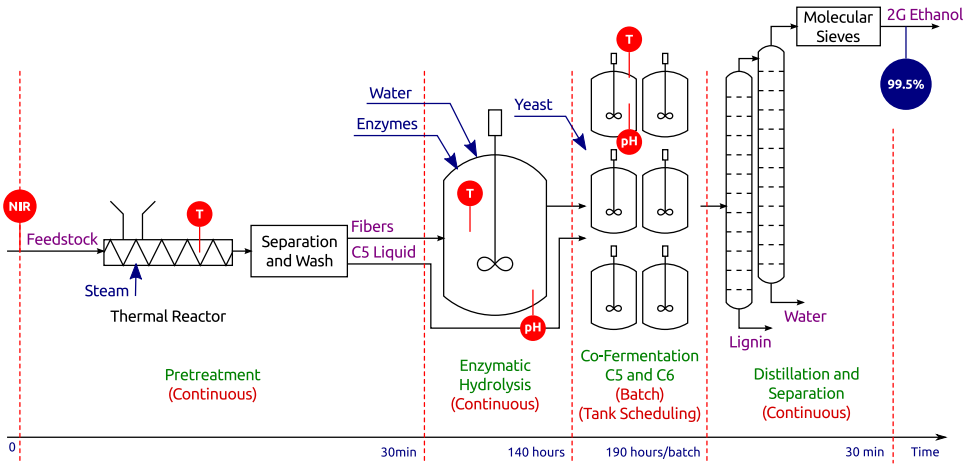
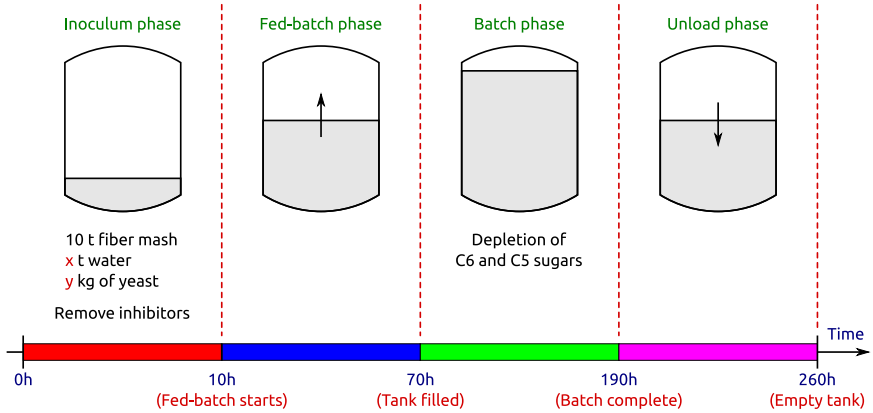


Figure C.1: Biorefinery diagram with assumed instrumentation. Pretreatment, the enzymatic hydrolysis, and purification are continuous processes, while fermentation occurs in scheduled batch reactors. Feedstock composition is assumed to be known, and can be measured in reality with NIR equipment.

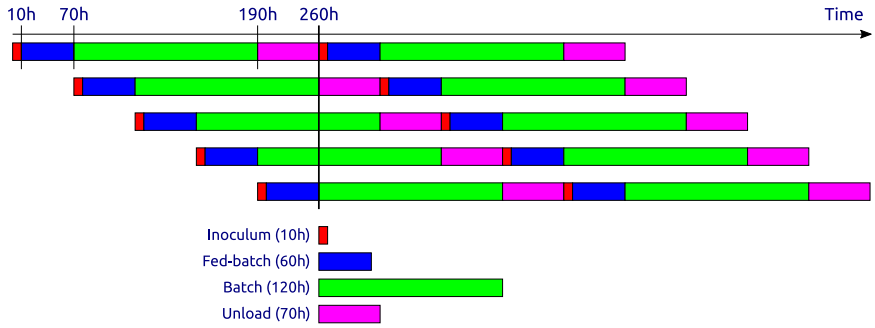
time of 140 h. The tanks are equipped with pH and temperature controllers in order to keep optimal conditions for the enzymatic activity: e.g. $pH = 5$, and $T = 50\,^{\circ}\mathrm{C}$ [58]. Enzymes are added by a pump from a storage tank. The enzyme dosage F_e can be adjusted accordingly and constitutes a degree of freedom in the optimization problem.

The fermentation process runs in scheduled batch mode in a number of tanks with a maximum holdup of $250\,\mathrm{m}^3$. The fermentation tanks have pH and temperature controllers. Each batch consists of an inoculum phase, the fed-batch phase (which cannot be neglected because it takes 60 h to fill the tank, time when reactions already take place), the batch stage, and an unload step. These stages are illustrated in Figure C.2(a). In the inoculum phase, 10 t of hydrolyzed fibers rich in glucose are mixed with M_y kilograms of yeast and diluted with water. The amount of yeast addition M_y is one of the degrees of freedom for operation. The fed-batch phase starts as soon as the inhibitors were removed, after about 10 h. The fermentation tank is filled up to 220 t with a constant feed rate calculated as the sum between the enzymatic hydrolysis outflow rate and the C5 liquid from the pretreatment process. Once the tank is filled, the batch phase begins where the C5 and C6 sugars are depleted. The batch stage has a fixed duration set to 120 h.

A large scale biorefinery has several fermentation reactors running in parallel following a certain scheduling algorithm. Figure C.2(b) shows the scheduling



(a) Fermentation process consisting of 4 phases: inoculum (10 h), fed-batch (60 h), batch (120 h), and unload (70 h).



(b) An example of a fermentation process with 5 scheduled reactors. The reactors are scheduled such that the liquefied fiber inflow and ethanol outflow stay constant with minimum interrupts.

Figure C.2: Fermentation process: sequential operation and scheduling.

algorithm for 5 reactors such that the inflow and outflow rates have minimum interruptions. This is achieved by aligning in series the fed-batch phases from all tanks, and by synchronizing the unload stages.

The distillation and purification phase separates lignin and water from ethanol. Lignin is then recovered as bio-pellets in a nearby evaporation unit, while ethanol achieves a high concentration of 99.5 % with the help of several molecular sieves.

One can change either the reaction time by modifying the retention time of each individual process, or adjusting the pretreatment temperature, enzyme dosage, and yeast seed to maximize ethanol yield. In large scale plants it is desired to keep a constant throughput. This constraint translates to a fixed pretreatment time, i.e. 15 min, a constant enzymatic hydrolysis time of 140 h, and a fermentation time of 190 h per batch for the demonstration scale plant studied here. The degrees of

freedom then become: the thermal reactor temperature T_{tr}, the enzyme dosage F_e, and yeast seed M_y, which are the key process parameters sought by the optimization layer.

C.2.2 The Optimization Layer

This study aims to develop an optimization method that can be implemented as a supervisory layer at a large scale facility. Figure C.3 shows the role of the optimization layer, and how it interacts with the system identification layer, the control system, and the real plant. The model identification layer utilizes real measurements to calibrate the plant model such that predictions become more accurate. The control layer translates the optimal setpoints into actuator commands to ensure reference tracking for the key process parameters. A pretreatment temperature regulator has been designed in a previous study [12]. The enzyme dosage controller is based on a flow rate meter and a variable positioning valve mounted on the supply pipe. The yeast seed controller requires a mass estimator for added yeast and an on/off valve. The control system is assumed to maximize the efficiency at a constant throughput, which is common practice for a new plant [81].

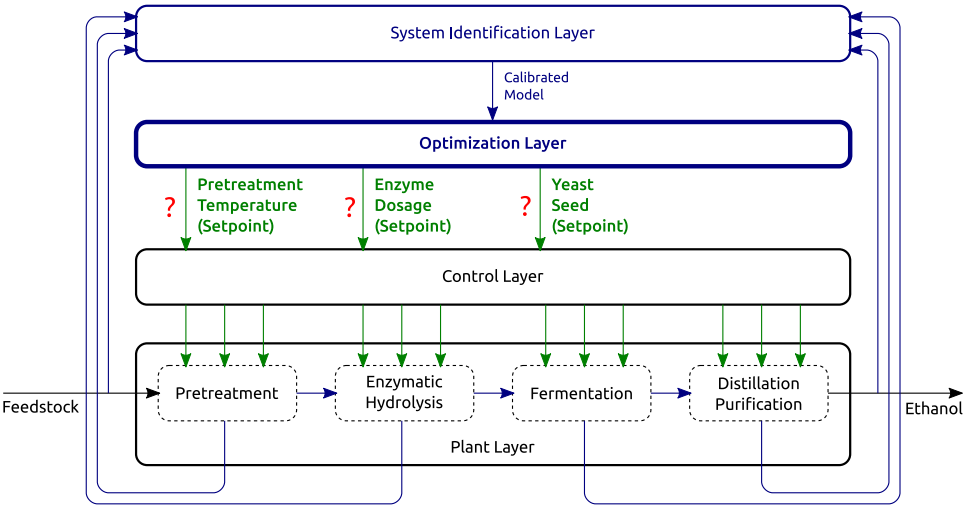


Figure C.3: Block diagram showing the interaction between the optimization layer designed in this study and the real plant. The optimization layer calculates setpoints for pretreatment temperature, enzyme dosage and yeast seed. All models are calibrated by the system identification layer based on plant measurements.

The optimization solver can be triggered either when prices change (e.g. ethanol

price increases, enzymes price decreases, yeast can be grown at a lower price etc.), or when disturbances occur in the process (e.g. changes in feedstock content due to biomass composition variety, or biomass type change). The solution of the optimization problem provides setpoints for the pretreatment temperature, enzyme dosage, and yeast seed. The system constraints are formulated based on validated large scale models for: hydrothermal pretreatment with steam [5], enzymatic hydrolysis [6], and C5 and C6 co-fermentation [74]. The optimizer searches for the best nominal operational point by evaluating different scenarios with the help of an integrated biorefinery simulator that embeds all these models. The optimization problem has the following generic formulation [82]:

$$\begin{aligned} \max_{u} \quad & c(x,u) \\ \text{subject to} \quad & f(x,u) = 0 \\ & g(x,u) \leq 0 \end{aligned} \tag{C.1}$$

where u is a vector of the decision variables or degrees of freedom, x represents the process variables, and $c(x,u)$ is the objective function. $f(x,u) = 0$ and $g(x,u) \leq 0$ are equality and inequality constraints as functions of process and decision variables. Optimization problem (C.1) is solved in a scientific software that has a solver for finding the minimum of a constrained nonlinear multivariable function.

C.2.3 Mathematical Models

The optimization layer uses a dynamic biorefinery simulator to calculate the stabilized or steady-state process outputs for pretreatment and enzymatic hydrolysis. The mathematical models are complex and nonlinear. Finding an analytical solution for steady-states is not trivial. An alternative is to run a sufficient long simulation until all outputs are stabilized. The fermentation model is a batch process and the outputs are collected after running a dynamic simulation for 190 h, i.e. the end of the batch phase, which is a fixed amount of time due to a constant throughput. The pretreatment and enzymatic hydrolysis models have already been published in [5] and [6], and a summary of all equations is given in the supplementary material. The fermentation model is detailed next.

C.2.3.1 C5 and C6 Co-Fermentation Model

The fermentation tank is modeled as a continuous stirred tank reactor (CSTR) with reaction kinetics derived from [74]. Since it can take 70 h for filling the tank, it is important to track the total mass change in time:

$$\frac{dM_f}{dt} = F_{in_f} - F_{out_f} \tag{C.2}$$

where M_f is the fermenter hold-up in kg, and F_{in_f} and F_{out_f} are the mass inflow and outflow rates in $\mathrm{kg\,s^{-1}}$.

Component composition is tracked with the following equation for variable hold-up:

$$\frac{d(C_f M_f)}{dt} = F_{in_f} C_{in_f} - F_{out_f} C_f + R_f M_f \tag{C.3}$$

where C_f is the composition vector in $\mathrm{g\,kg^{-1}}$, C_{in_f} is the inflow composition of liquefied fibers coming from the enzymatic hydrolysis process, and R_f is the reaction rate vector in $\mathrm{g/(kgs)}$.

Substituting Equation (C.2) into (C.3) yields:

$$\frac{dC_f}{dt} = \frac{1}{M_f}\left[F_{in_f}\left(C_{in_f} - C_f\right)\right] + R_f \tag{C.4}$$

The reaction rates are calculated using standard global black box model to represent fermentation activity [119]:

- Glucose uptake:

$$R_G = -q_{G_I} \tag{C.5}$$

  where R_G is the glucose uptake rate, and equals the glucose consumption rate with inhibition q_{G_I}, which is calculated as follows:

$$q_{G_I} = \frac{1}{Y_{Eth_G}} q_{Eth_{G_I}} \tag{C.6}$$

  Y_{Eth_G} is the yield parameter for ethanol production from glucose. $q_{Eth_{G_I}}$ represents ethanol production rate from glucose with inhibition:

$$q_{Eth_{G_I}} = q_{Eth_G} \cdot I_{Eth_G} \cdot I_{F_G} \cdot I_{A_G} \cdot I_{HMF_G} \tag{C.7}$$

  q_{Eth_G} is ethanol production from glucose only with substrate inhibition, while I_{Eth_G}, I_{F_G}, I_{A_G}, and I_{HMF_G} are inhibitory terms for ethanol (product inhibition), furfural, acetate, and 5-HMF respectively. q_{Eth_G} is modeled as in [120] with pH dependency extension:

$$q_{Eth_G} = q_{Max_G}(pH) \cdot C_{Cell} \frac{C_G}{K_{SP_G} + C_G + \dfrac{{C_G}^2}{K_{IP_G}}} \tag{C.8}$$

  where C_G and C_{Cell} are glucose and cell biomass concentrations, and K_{SP_G} and K_{IP_G} are glucose self-inhibition terms.

The pH dependency is built empirically to resemble a bell with q_{Max_G} as peak:

$$q_{Max_G}(pH) = q_{Max_G} \frac{K_{0_G}}{\left(1 + \frac{10^{pH}}{K_{1_G}} + \frac{K_{2_G}}{10^{pH}}\right)} \tag{C.9}$$

where K_{0_G}, K_{1_G}, and K_{2_G} parametrize the shape of the bell.

Ethanol inhibition on glucose uptake or product inhibition has been defined in [120]:

$$I_{Eth_G} = 1 - \left(\frac{C_{Eth}}{P_{MP_G}}\right)^{\gamma_G} \tag{C.10}$$

with P_{MP_G} and γ_G as parameters. C_{Eth} is the concentration of ethanol.

Furfural, acetate and 5-HMF inhibition are modeled as below:

$$I_{i_G} = \frac{K_{I_{i_G}}}{K_{I_{i_G}} + C_i} \tag{C.11}$$

where I_{i_G} is the inhibition from component i, $K_{I_{i_G}}$ is the inhibitory constant, and C_i is the concentration. Index i can be $\{_{F,\ Ac,\ HMF}\}$ signifying furfural, acetate, or 5-HMF.

- Xylose uptake follows the same equation structure as glucose uptake but with different parameter values:

$$R_X = -q_{X_I} \tag{C.12}$$

Xylose consumption with inhibition is calculated based on ethanol production from xylose $q_{Eth_{X_I}}$ with inhibition, and yield parameter Y_{Eth_X}:

$$q_{X_I} = \frac{1}{Y_{Eth_X}} q_{Eth_{X_I}} \tag{C.13}$$

Ethanol production from xylose comprises inhibition terms from product I_{Eth_X}, furfural I_{F_X}, acetate I_{Ac_X}, and 5-HMF I_{HMF_X}:

$$q_{Eth_{X_I}} = q_{Eth_X} \cdot I_{Eth_X} \cdot I_{F_X} \cdot I_{Ac_X} \cdot I_{HMF_X} \tag{C.14}$$

q_{Eth_X} shows ethanol production with substrate inhibition modeled as in [120] with pH dependency extension:

$$q_{Eth_X} = q_{Max_X}(pH) \cdot C_{Cell} \frac{C_X}{K_{SP_X} + C_X + \frac{{C_X}^2}{K_{IP_X}}} \tag{C.15}$$

C_X and C_{Cell} are xylose and cell biomass concentrations, while K_{SP_X} and K_{IP_X} are xylose inhibitory parameters.

The pH dependency curve is parametrized in K_{0_X}, K_{1_X}, and K_{2_X} with q_{Max_X} as peak:

$$q_{Max_X}(pH) = q_{Max_X} \frac{K_{0_X}}{\left(1 + \frac{10^{pH}}{K_{1_X}} + \frac{K_{2_X}}{10^{pH}}\right)} \tag{C.16}$$

Product inhibition is similar to the glucose case with parameters for xylose P_{MP_X} and γ_X:

$$I_{Eth_X} = 1 - \left(\frac{C_{Eth}}{P_{MP_X}}\right)^{\gamma_X} \tag{C.17}$$

Furfural, acetate and 5-HMF inhibition have the following equations:

$$I_{i_X} = \frac{K_{I_{i_X}}}{K_{I_{i_X}} + C_i} \tag{C.18}$$

with index $i \in \{_F,\ _{Ac},\ _{HMF}\}$.

- Ethanol production is the sum between the rates from glucose and xylose uptake shown in Equations (C.6) and (C.13):

$$R_{Eth} = q_{G_I} + q_{X_I} \tag{C.19}$$

- Furfural uptake rate R_F is calculated as:

$$R_F = -q_F \tag{C.20}$$

$$q_F = q_{Max_F} C_{Cell} \frac{C_F}{K_{F_S} + C_F} \tag{C.21}$$

where q_{Max_F} is the maximum uptake rate, C_{Cell} is the cell biomass concentration and K_{F_S} is a substrate inhibition parameter.

- 5-HMF uptake is modeled similarly as in the furfural case with an inhibitory term addition due to furfural:

$$R_{HMF} = -q_{HMF_I} \tag{C.22}$$

$$q_{HMF_I} = q_{HMF} \cdot I_{F_{HMF}} \tag{C.23}$$

$$q_{HMF} = q_{Max_{HMF}} C_{Cell} \frac{C_{HMF}}{C_{HMF} + K_{HMF_S}} \tag{C.24}$$

$$I_{F_{HMF}} = \frac{K_{I_{HMF_F}}}{K_{I_{HMF_F}} + C_F} \tag{C.25}$$

where K_{HMF_S} and $K_{I_{HMF_F}}$ are inhibition parameters.

- Acetate uptake and production:

$$R_{Ac} = q_{Ac_{HMF}} - q_{Ac} \tag{C.26}$$

where $q_{Ac_{HMF}}$ is the production rate and q_{Ac} is the uptake rate. The acetate production rate is calculated based on Equation (C.23) and the yield parameter $Y_{Ac_{HMF}}$:

$$q_{Ac_{HMF}} = q_{HMF_I} \cdot Y_{Ac_{HMF}} \tag{C.27}$$

Acetate uptake is modeled similarly to 5-HMF uptake:

$$q_{Ac} = q_{Max_{Ac}} C_{Cell} \frac{C_{Ac}}{C_{Ac} + K_{Ac_S}} \tag{C.28}$$

with $q_{Max_{Ac}}$ as maximum uptake rate and K_{Ac_S} as substrate inhibition term.

- CO_2 production occurs in glucose, xylose and acetate uptake:

$$R_C = q_{G_I} Y_{C_G} + q_{X_I} Y_{C_X} + q_{Ac} Y_{C_{Ac}} \tag{C.29}$$

where Y_{C_G}, Y_{C_X} and $Y_{C_{Ac}}$ are yield parameters.

- Cell biomass growth is modeled as in [120]:

$$R_{Cell} = \mu_T \tag{C.30}$$

where μ_T is the specific growth of C5 and C6 mixture detailed as follows:

$$\mu_T = \frac{C_G}{C_G + C_X} \mu_G + \frac{C_X}{C_G + C_X} \mu_X \tag{C.31}$$

μ_G and μ_X are the specific growths on glucose and xylose:

$$\mu_G = (q_{G_I} - m_G C_{Cell}) Y_{Cell_G} \tag{C.32}$$

$$\mu_X = (q_{X_I} - m_X C_{Cell}) Y_{Cell_X} \tag{C.33}$$

m_G and m_X are maintenance coefficients for glucose and xylose, while Y_{Cell_G} and Y_{Cell_X} are yield parameters.

C.2.3.2 Biorefinery Model Summary

Table C.1 offers a summary of the integrated model complexity. The overall model accounts for 96 kinetic parameters, 580 states, 10 inputs and 25 outputs. The table also offers a split of the modular model based on the refinery step. The high number of states in pretreatment and enzymatic hydrolysis is due to the computational fluid dynamics tools (the convection equation discretized in space) used for modeling the thermal reactor and the first enzymatic hydrolysis tank. Nominal values for kinetic and feed parameters are given in Table C.3 from the model initialization section.

Table C.1: Model summary: number of parameters, states, inputs and outputs. Half of the pretreatment outputs (the pretreated fibers) are directed to the enzymatic hydrolysis process, while the other half (the C5 liquid) is connected to fermentation. The outputs from the enzymatic hydrolysis are connected to fermentation.

Model	Parameters	States	Inputs	Outputs
Pretreatment	17	298	10	36
Enzymatic Hydrolysis	46	257	18	19
Fermentation (1 tank)	33	25	37	25
Total	96	580	10	25

C.2.4 Plantwide Optimization Methodology

The methodology steps for finding the optimal operational point of a plant are extended from [82]:

1. Select the objective or cost function;
2. Identify the decision variables;
3. Formulate process model constraints and set bounds for decision variables;
4. Formulate and solve the optimization problem;
5. Sensitivity and uncertainty analysis of the optimal solution.

The optimal solution is analyzed from a sensitivity point of view using similar tools as in [5, 9, 66]. Mathematical models that describe complex systems are often over-parametrized. The sensitivity analysis quantifies the relation between the cost function and model parameters when the system runs at the optimal point. The aim is to rank all model parameters by their significance with respect to the profit

value at the nominal operational point. Also a subset of relevant parameters can be extracted for calculating the uncertainty bounds.

A measure of sensitivity suitable for steady-state signals is [8, 64]:

$$\delta_k = \frac{\partial c_k}{\partial \theta_k} \frac{\theta_k}{c_{ss_k}} \tag{C.34}$$

where $\partial c_k / \partial \theta_k$ is the variation in profit with respect to a model parameter, and is calculated based on finite differences. θ_k is the k^{th} parameter, and c_{ss_k} is the profit or the value of the cost function in steady-state. All model parameters are sorted in descending order with respect to δ_k, and a subset is created with all parameters that have δ_k above a threshold. The reduced subset of model parameters is then used in the uncertainty analysis.

The propagation of uncertainty is analyzed with a Monte Carlo procedure as described in [9]:

1. Define input uncertainty;
2. Parameter sampling;
3. Monte Carlo simulations;
4. Output uncertainty.

The input uncertainty is defined with standard deviations and correlation matrices obtained from previous studies. Dealing with many parameters implies a large number of combinations of parameter values with high correlation between them. In order to reduce the number of parameter samples, a Latin Hypercube Sampling (LHS) technique with correlation control is utilized [65]. LHS generates less samples of parameters but still statistically plausible with the help of a distribution function, standard deviation, and correlation matrix. For each set of samples, a simulation is then run and the output is collected. After all Monte-Carlo simulations are performed, enough output information is obtained to statistically compute the median and the 5th-95th percentile confidence interval.

C.3 Results and Discussion

This section starts with model initialization where the values of both model and feed parameters are given in tabular format. The solution of the optimization problem is then found and discussed. The sensitivity analysis with respect to model kinetic parameters follows. A parameter subset is then identified, which is used in Monte Carlo simulations for quantifying the uncertainty of the optimal solution. The

costs and profit curves are also computed with uncertainty bounds. Uncertainty can be embedded in the formulation of the optimal problem and the results are shown for a stochastic optimization problem. The section ends with dynamic simulations showing the refinery operation at the optimal point.

C.3.1 Model Initialization

Table C.2 shows the feed parameters, i.e. raw biomass inflow rate, composition and initial temperature. The inflow rate is set to $1000\,\mathrm{kg\,h^{-1}}$, the throughput of a demonstration scale plant. The biomass composition resembles wheat straw with an initial dry matter of 89 % [40].

Table C.2: Biorefinery inputs: inflow rate, raw biomass composition, and initial temperature.

	Description	Value	Unit	% of dry matter
1	Inflow rate	1000	$\mathrm{kg\,h^{-1}}$	
2	Cellulose	360	$\mathrm{g\,kg^{-1}}$	40.45
3	Xylan	187	$\mathrm{g\,kg^{-1}}$	21.01
4	Arabinan	23	$\mathrm{g\,kg^{-1}}$	2.58
5	Lignin	200	$\mathrm{g\,kg^{-1}}$	22.47
6	Acetyls	44	$\mathrm{g\,kg^{-1}}$	4.94
7	Ash	26	$\mathrm{g\,kg^{-1}}$	2.92
8	Water	110	$\mathrm{g\,kg^{-1}}$	-
9	Other	50	$\mathrm{g\,kg^{-1}}$	5.63
10	Temperature	15	°C	

Table C.3 indicates the values with units for all 96 model parameters. The table is split into pretreatment, enzymatic hydrolysis and fermentation. The model parameter values are taken from [6], [5], and [74].

Table C.3: Model kinetic parameters: description, numerical values and measuring unit.

	θ_k	Description	Value	Unit
	Pretreatment			
1	E_{Xo}	Xylooligomers activation energy	298 011	$\mathrm{J\,mol^{-1}}$
2	E_X	Xylose activation energy	304 680	$\mathrm{J\,mol^{-1}}$
3	E_G	Glucose activation energy	335 614	$\mathrm{J\,mol^{-1}}$
4	E_{PL}	Pseudo-lignin activation energy	325 629	$\mathrm{J\,mol^{-1}}$
5	E_F	Furfural activation energy	327 253	$\mathrm{J\,mol^{-1}}$
6	E_{HMF}	5-HMF activation energy	300 000	$\mathrm{J\,mol^{-1}}$

7	E_{Ac}	Organic acid activation energy	242 687	$\mathrm{J\,mol^{-1}}$
8	E_A	Arabinose activation energy	61 229	$\mathrm{J\,mol^{-1}}$
9	k_{Xo}	Xylooligomers reaction constant	2.78×10^{31}	$\mathrm{s^{-1}}$
10	k_X	Xylose reaction constant	1.31×10^{34}	$\mathrm{s^{-1}}$
11	k_G	Glucose reaction constant	1.11×10^{35}	$\mathrm{s^{-1}}$
12	k_{PL}	Pseudo-lignin reaction constant	1.03×10^{33}	$\mathrm{s^{-1}}$
13	k_F	Furfural reaction constant	5.09×10^{33}	$\mathrm{s^{-1}}$
14	k_{HMF}	5-HMF reaction constant	1×10^{31}	$\mathrm{s^{-1}}$
15	k_{Ac}	Organic acid reaction constant	4.88×10^{24}	$\mathrm{s^{-1}}$
16	k_A	Arabinose reaction constant	106 225	$\mathrm{s^{-1}}$
17	α_{PL}	Pseudo-lignin and inhibitors fraction	0.1019	–
	Enzymatic Hydrolysis			
18	α_C^E	Endo-exo cellulase fraction	0.25	–
19	α_C^G	β-glucosidase fraction	0.25	–
20	α_X^E	Endo-exo xylanase fraction	0.25	–
21	α_X^G	β-xylosidase fraction	0.25	–
22	K_1	Cellulose to cellobiose reaction constant	0.005 916	$\mathrm{kg/(gs)}$
23	K_2	Cellulose to glucose reaction constant	0.006 507 5	$\mathrm{kg/(gs)}$
24	K_3	Cellobiose to glucose reaction constant	0.005 522 7	$\mathrm{kg/(gs)}$
25	K_4	Xylan to xylooligomers reaction constant	0.002 002 6	$\mathrm{kg/(gs)}$
26	K_5	Xylan to xylose reaction constant	0.003 393 6	$\mathrm{kg/(gs)}$
27	K_6	Xylooligomers to xylose reaction constant	0.002 822 8	$\mathrm{kg/(gs)}$
28	K_7	Enzyme deactivation reaction constant	2.5×10^{-7}	$\mathrm{kg/(gs)}$
29	$K_{A_C}^E$	Exo-endo cellulase Langmuir adsorption constant	1.0444	–
30	$K_{A_X}^E$	Exo-endo xylanase Langmuir adsorption constant	0.378 44	–
31	$K_{A_C}^G$	β-glucosidase Langmuir adsorption constant	0.056 976	–
32	$K_{A_X}^X$	Xilosidase Langmuir adsorption constant	0.093 253	–
33	$E_{M_C}^E$	Maximum exo-endo cellulase adsorption	0.016 042	–
34	$E_{M_C}^G$	Maximum β-glucosidase adsorption	1.5×10^{-5}	–
35	$E_{M_X}^E$	Maximum endo-exo xylanase adsorption constant	0.389 78	–
36	$E_{M_X}^X$	Maximum xylosidase adsorption	0.511 78	–
37	I_{C_1}	Cellobiose inhibition on r_1	0.020 14	$\mathrm{g\,kg^{-1}}$
38	I_{G_1}	Glucose inhibition on r_1	0.102 55	$\mathrm{g\,kg^{-1}}$
39	I_{Xo_1}	Xylooligomers inhibition on r_1	0.007 814 5	$\mathrm{g\,kg^{-1}}$
40	I_{X_1}	Xylose inhibition on r_1	0.015 03	$\mathrm{g\,kg^{-1}}$
41	I_{Eth_1}	Ethanol inhibition on r_1	0.15	$\mathrm{g\,kg^{-1}}$
42	I_{C_2}	Cellobiose inhibition on r_2	69.539	$\mathrm{g\,kg^{-1}}$

43	I_{G_2}	Glucose inhibition on r_2	0.067554	$\mathrm{g\,kg^{-1}}$
44	I_{Xo_2}	Xylooligomers inhibition on r_2	0.059612	$\mathrm{g\,kg^{-1}}$
45	I_{X_2}	Xylose inhibition on r_2	0.14843	$\mathrm{g\,kg^{-1}}$
46	I_{G_3}	Glucose inhibition on r_3	8.7211	$\mathrm{g\,kg^{-1}}$
47	I_{Xo_3}	Xylooligomers inhibition on r_3	111.6822	$\mathrm{g\,kg^{-1}}$
48	I_{X_3}	Xylose inhibition on r_3	210.1911	$\mathrm{g\,kg^{-1}}$
49	I_{O_3}	Overall inhibition on r_3	15.949	$\mathrm{g\,kg^{-1}}$
50	I_{C_4}	Cellobiose inhibition on r_4	53.4804	$\mathrm{g\,kg^{-1}}$
51	I_{G_4}	Glucose inhibition on r_4	2.0899	$\mathrm{g\,kg^{-1}}$
52	I_{Xo_4}	Xylooligomers inhibition on r_4	113.4492	$\mathrm{g\,kg^{-1}}$
53	I_{X_4}	Xylose inhibition on r_4	233.0874	$\mathrm{g\,kg^{-1}}$
54	I_{C_5}	Cellobiose inhibition on r_5	2.7413	$\mathrm{g\,kg^{-1}}$
55	I_{G_5}	Glucose inhibition on r_5	4.7951	$\mathrm{g\,kg^{-1}}$
56	I_{Xo_5}	Xylooligomers inhibition on r_5	83.5479	$\mathrm{g\,kg^{-1}}$
57	I_{X_5}	Xylose inhibition on r_5	271.2334	$\mathrm{g\,kg^{-1}}$
58	I_{C_6}	Cellobiose inhibition on r_6	46.9663	$\mathrm{g\,kg^{-1}}$
59	I_{G_6}	Glucose inhibition on r_6	3.0412	$\mathrm{g\,kg^{-1}}$
60	I_{X_6}	Xylose inhibition on r_6	198.3351	$\mathrm{g\,kg^{-1}}$
61	I_{O_6}	Overall inhibition on r_6	28.2079	$\mathrm{g\,kg^{-1}}$
62	R_A	Severity dependency	9	–
63	R_B	Severity dependency	2.915	–
	Fermentation			
64	$Y_{CO_{2_G}}$	CO_2 production from glucose uptake	0.47	$\mathrm{kg\,kg^{-1}}$
65	$Y_{CO_{2_X}}$	CO_2 production from xylose uptake	0.4	$\mathrm{kg\,kg^{-1}}$
66	K_{F_S}	Furfural uptake self inhibition constant	0.05	$\mathrm{g\,kg^{-1}}$
67	$K_{I_{F_G}}$	Glucose inhibition on furfural uptake	0.75	$\mathrm{g\,kg^{-1}}$
68	$K_{I_{HMF_F}}$	Furfural inhibition on 5-HMF uptake	0.25	$\mathrm{g\,kg^{-1}}$
69	$K_{I_{F_X}}$	Xylose inhibition on furfural uptake	0.35	$\mathrm{g\,kg^{-1}}$
70	$q_{F_{Max}}$	Maximum furfural uptake	4.6706×10^{-5}	$\mathrm{s^{-1}}$
71	K_{IP_G}	Glucose uptake self inhibition parameter	4890	$\mathrm{g\,kg^{-1}}$
72	K_{SP_G}	Glucose uptake self inhibition parameter	1.342	$\mathrm{g\,kg^{-1}}$
73	P_{MP_G}	Ethanol inhibition on glucose uptake	103	$\mathrm{g\,kg^{-1}}$
74	γ_G	Ethanol inhibition on glucose uptake	1.42	–
75	Y_{Eth_G}	Ethanol production from glucose uptake	0.47	$\mathrm{kg\,kg^{-1}}$
76	Y_{Cell_G}	Biomass growth on glucose	0.115	$\mathrm{kg\,kg^{-1}}$
77	m_G	Maintenance coefficient for biomass growth on glucose	2.6944×10^{-5}	$\mathrm{s^{-1}}$
78	q_{Max_G}	Maximum glucose uptake rate	0.000318	$\mathrm{s^{-1}}$
79	K_{IP_X}	Xylose uptake self inhibition parameter	81.3	$\mathrm{g\,kg^{-1}}$
80	K_{SP_X}	Xylose uptake self inhibition parameter	3.4	$\mathrm{g\,kg^{-1}}$

81	P_{MP_X}	Ethanol inhibition on xylose uptake	100.2	$\mathrm{g\,kg^{-1}}$
82	γ_X	Ethanol inhibition on xylose uptake	0.608	–
83	Y_{Eth_X}	Ethanol production from xylose uptake	0.4	$\mathrm{kg\,kg^{-1}}$
84	Y_{Cell_X}	Biomass growth on xylose	0.162	$\mathrm{kg\,kg^{-1}}$
85	m_X	Maintenance coefficient for biomass growth on xylose	1.8611×10^{-5}	$\mathrm{s^{-1}}$
86	q_{Max_X}	Maximum xylose uptake rate	0.00083444	$\mathrm{s^{-1}}$
87	K_{Ac_S}	Acetate uptake self inhibition	2.5	$\mathrm{g\,kg^{-1}}$
88	$K_{I_{Ac_G}}$	Acetate inhibition on glucose uptake	2.74	$\mathrm{g\,kg^{-1}}$
89	$K_{I_{Ac_X}}$	Acetate inhibition on xylose uptake	0.2	$\mathrm{g\,kg^{-1}}$
90	$Y_{Ac_{HMF}}$	Acetate production from 5-HMF uptake	0.23392	$\mathrm{kg\,kg^{-1}}$
91	$Y_{CO_{2_S}}$	CO_2 production from 5-HMF uptake	0.1	$\mathrm{kg\,kg^{-1}}$
92	$q_{Max_{Ac}}$	Maximum acetate uptake rate	1.2292×10^{-5}	$\mathrm{s^{-1}}$
93	K_{HMF_S}	5-HMF uptake self inhibition	0.5	$\mathrm{g\,kg^{-1}}$
94	$K_{I_{HMF_G}}$	5-HMF inhibition on glucose uptake	2	$\mathrm{g\,kg^{-1}}$
95	$K_{I_{HMF_X}}$	5-HMF inhibition on xylose uptake	10	$\mathrm{g\,kg^{-1}}$
96	$q_{Max_{HMF}}$	Maximum 5-HMF uptake rate	8.7576×10^{-5}	$\mathrm{s^{-1}}$

C.3.2 The Optimization Problem

The generic method steps from [82] are detailed next:

1. Select the objective or cost function:

   The cost function from this study represents the profit for one fermentation batch defined as the difference between ethanol revenue and operating costs related to biomass, steam, enzymes, and yeast:

   $$c(M_{Eth}, F_b, F_s, F_e, M_y) = M_{Eth}(t_f)P_{Eth} - (F_bP_b + F_sP_s + F_eP_e + M_yP_y) \tag{C.35}$$

   Ethanol revenue is calculated as $M_{Eth}(t_f)P_{Eth}$, i.e. mass of ethanol in kg at the end of the batch phase t_f times its price per kilogram P_{Eth}. M_y is the mass of yeast added to the fermentation tank in the inoculum phase. The operating costs are defined as flow rate or mass of utility times its price. The refinery consists of two continuous processes, i.e. pretreatment and enzymatic hydrolysis, and a batch process, i.e. fermentation. The weights P_b (cost of biomass), P_s (cost of steam), and P_e (cost of enzymes) are related to the continuous processes, i.e. pretreatment and liquefaction, and are measured in unitcost/(kg/h). P_y (cost of yeast) is measured in unitcost/kg as it gets multiplied with a mass. The overall measuring unit of the cost function becomes the unitcost, which can represent any currency.

The feedstock flow rate or refinery throughput F_b is kept constant and therefore can be dropped from the cost function as the derivative of the cost function with respect to F_b becomes 0. Table C.4 shows the weight values used in this study. Distillation cost is a function of water content of the stream from fermentation. As none of the decision variables influence the water content, the distillation costs can also be removed from the objective function, F_s showing only the steam used for pretreatment. The cost function then becomes:

$$c(M_{Eth}, F_s, F_e, M_y) = M_{Eth}(t_f)P_{Eth} - (F_sP_s + F_eP_e + M_yP_y) \tag{C.36}$$

Table C.4: Cost function weights (prices).

Parameter	Description	Value
P_{Eth}	Ethanol	5 unitcost/kg
P_e	Enzymes	25 unitcost/(kg/h)
P_s	Steam	1 unitcost/(kg/h)
P_y	Yeast	50 unitcost/kg

2. Identify the decision variables:

   The outcome of the pretreatment process is sensitive to the thermal reactor temperature and retention time [50]. However, it is desired to have a constant throughput as mentioned earlier, which makes the retention time constant. The thermal reactor temperature then becomes the only degree of freedom in pretreatment.

   The key parameters in enzymatic hydrolysis are: pH, temperature, and concentration of enzymes. The enzymatic activity is a function of pH and temperature, which resemble Gaussian curves with single peaks at pH of 5 units and temperature 50 °C [6]. Any deviations from the optimal point would reduce the enzymatic efficiency. Control loops keep the pH and temperature close to optimality [11] and it is not indicated to vary these variables. However, the concentration of enzymes can be adjusted by changing the inflow rate of enzymes F_e and constitutes the only degree of freedom in enzymatic hydrolysis for the optimization problem.

   The efficiency of the fermentation process is a function of pH, temperature, and yeast seed. The optimal pH level of the GMO yeast is relatively close to that of the enzymes, i.e. 5.5 units. The optimal fermentation temperature is at 35 °C, which is different than the enzymatic optimal temperature. Controllers keep the pH and temperature conditions at the GMO yeast optimal levels throughout

the entire fermentation process. The only degree of freedom considered for the optimization problem is the yeast seed M_y in the inoculum phase.

As a summary, the decision variables are: the pretreatment temperature T_{tr} defined as the set-point for the thermal reactor temperature controller, the inflow rate of enzymes F_e expressed in $\mathrm{kg\,h^{-1}}$, and the yeast seed M_y in kg as a set-point for the amount of yeast used to start the inoculum phase:

$$u = [T_{tr} \quad F_e \quad M_y]^\top \tag{C.37}$$

3. Process model constraints, and bounds for decision variables:

The dynamic integrated models for pretreatment and enzymatic hydrolysis are formulated as:

$$\dot{x} = f(x,u) \tag{C.38}$$

where $f(x,u)$ is a nonlinear function of states x and inputs u. The steady states are then found as the solution of $\dot{x} = 0$:

$$0 = f(x,u) \tag{C.39}$$

Due to the model complexity and the nonlinear nature of $f(x,u)$, an analytical solution to (C.39) cannot be easily found. As an alternative, the steady states are calculated by running a sufficient long simulation till all states stabilize.

The dynamic model for fermentation is described by:

$$\dot{x}_f = h(x_f, u_f) \tag{C.40}$$

where $h(x_f,u_f)$ represents a nonlinear complex model of states x_f and inputs u_f. The final states at time 190 h are found by integrating the model numerically (dynamic simulation):

$$x_f(t_f) = \int_0^{t_f} h(x_f,u_f)dt \tag{C.41}$$

where $t_f = 190\,\mathrm{h}$, i.e. the end of the batch phase.

The decision variables are bounded as follows:

$$\begin{array}{rcccl} 150 & \leq & T_{tr} & \leq & 210\,^\circ\mathrm{C} \\ 10 & \leq & F_e & \leq & 1000\,\mathrm{kg\,h^{-1}} \\ 10 & \leq & M_y & \leq & 1000\,\mathrm{kg} \end{array} \tag{C.42}$$

which allows a wide range of operation for searching the optimal point.

4. Formulate and solve the overall optimization problem:

$$\begin{aligned} \max_{T_{tr},F_e,M_y} \quad & M_{Eth}(t_f)P_{Eth} - (F_sP_s + F_eP_e + M_yP_y) \\ \text{subject to} \quad & 0 = f(x(t),u(t)) \\ & \dot{x}_f = h(x_f,u_f) \\ 150 \;&\leq\; T_{tr} \;\leq\; 210\,^{\circ}\text{C} \\ 10 \;&\leq\; F_e \;\leq\; 1000\,\text{kg}\,\text{h}^{-1} \\ 10 \;&\leq\; M_y \;\leq\; 1000\,\text{kg} \end{aligned} \tag{C.43}$$

The previous optimization problem is solved numerically leading to the following solution:

$$z_o = \begin{bmatrix} T_{tr} \\ F_e \\ M_y \end{bmatrix} = \begin{bmatrix} 172\,^{\circ}\text{C} \\ 110\,\text{kg}\,\text{h}^{-1} \\ 142\,\text{kg} \end{bmatrix} \tag{C.44}$$

The thermal reactor temperature should be set to 172 °C, the enzyme dosage is of about 109 $\text{kg}\,\text{h}^{-1}$, and the yeast seed is of 142 kg. This optimal set point gives a profit of:

$$c_o = 7.6714 \times 10^4\,\text{unitprofit} \tag{C.45}$$

disregarding raw biomass and distillation costs.

In order to gain process insight and to observe how pretreatment conditions affect the downstream processes, an iteration is created through pretreatment temperatures between 160 °C to 180 °C with a step of 1 °C. Each pretreatment temperature changes the fibers and C5 liquid composition. The enzymatic hydrolysis and fermentation processes are then optimized in an integrated manner for each pretreatment temperature:

$$\begin{aligned} \max_{F_e,M_y} \quad & M_{Eth}P_{Eth} - (F_eP_e + M_yP_y) \\ \text{subject to} \quad & 0 = f(x(t),u(t)) \\ & \dot{x}_f = h(x_f,u_f) \\ 10 \;&\leq\; F_e \;\leq\; 1000\,\text{kg}\,\text{h}^{-1} \\ 10 \;&\leq\; M_y \;\leq\; 1000\,\text{kg} \end{aligned} \tag{C.46}$$

In this way the pretreatment, liquefaction and fermentation costs, as well as refinery profit can be observed with respect to pretreatment conditions. The same methodology can be applied even if there are recycles between fermentation and liquefaction because these two processes are analyzed in an integrated manner in optimization problem (C.46). Algorithm 5 shows how to calculate the curves for

profit, costs, and optimal solution as functions of pretreatment temperature. z is the optimal solution returned by the optimization problem solver, c_p, c_{eh}, c_f are the pretreatment, enzymatic hydrolysis, and fermentation costs, and c is the value of the cost function or the profit.

Algorithm 5 Calculate optimal costs and profit with respect to pretreatment temperature given a fixed set of model parameters θ and feedstock composition C_b.

function $[z, c_p, c_{eh}, c_f, c] = \text{COSTS}(\theta, C_b)$
  Set a range of pretreatment temperatures $T_{tr} \leftarrow 160\,^\circ\text{C} : 1\,^\circ\text{C} : 180\,^\circ\text{C}$
  Set initial solution guess to $z_0 \leftarrow [100\,\text{kg}\,\text{h}^{-1} \quad 80\,\text{kg}]^\top$
  for Each temperature in T_{tr} **do**
    Run pretreatment process at temperature T_{tr_i} and obtain composition of pretreated fibers and C5 liquid, and steam inflow into the thermal reactor F_{s_i}.
    $z_i \leftarrow$ Solution of optimization problem (C.46) given the pretreated fibers composition and C5 liquid from previous step as inputs. Use as initial guess the solution from previous iteration z_{i-1}.
    Calculate mass of ethanol at final fermentation time: $M_{Eth_i} \leftarrow M(t_f) \cdot C_{Eth}(t_f)$ where t_f is the final batch time, $M(t_f)$ is the reactor mass in kg at time t_f, and $C_{Eth}(t_f)$ is the ethanol concentration at time t_f in $\text{g}\,\text{kg}^{-1}$.
    Enzyme dosage: $F_{e_i} \leftarrow z_i(1)$.
    Yeast seed: $M_{y_i} \leftarrow z_i(2)$.
    Calculate pretreatment cost: $c_{p_i} \leftarrow F_{s_i} P_s$.
    Calculate liquefaction cost: $c_{eh_i} \leftarrow F_{e_i} P_e$.
    Calculate fermentation cost: $c_{f_i} \leftarrow M_{y_i} P_y$.
    Calculate revenue: $r_i \leftarrow M_{Eth_i} P_{Eth}$.
    Calculate profit $c_i \leftarrow r_i - (c_{p_i} + c_{eh_i} + c_{f_i})$.

C.3.3 Sensitivity Analysis of Profit Value at the Optimal Point

Figure C.4 illustrates the sensitivity analysis of the profit curve with respect to all model parameters from Table C.3 calculated at the optimal solution (C.44).

The most sensitive parameter is E_F, i.e. the furfural formation activation energy. Furfural is a strong fermentation inhibitor produced during pretreatment, and ethanol yield is directly affected by the amount of furfural. The next sensitive parameter is Y_{PS_G}, a yield parameter indicating the amount of ethanol in g produced per 1 g of glucose. E_G or glucose activation energy follows indicating that cellulose degradation in pretreatment impacts the ethanol yield. Three more fermentation parameters with similar sensitivity follow, i.e. maximum acetate uptake parameter

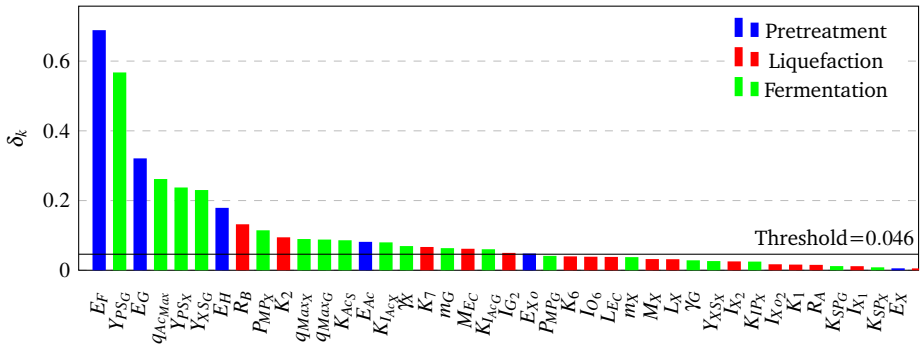


Figure C.4: Sensitivity measure δ_k of profit value with respect to model parameters.

$q_{Ac_{Max}}$, cell biomass yield on glucose Y_{XS_G}, and ethanol yield on xylose Y_{PS_X}. 5-HMF production during pretreatment has a relatively high sensitivity too as it influences both the glucose yield, by degrading it further, and also by inhibiting ethanol production in fermentation.

The first sensitive enzymatic hydrolysis parameter is R_B, i.e. the severity dependence of the enzymatic activity. R_B shows the importance of biomatrix opening from the pretreatment process as a structural breakdown of the fiber, which affects cellulose accessibility for enzymes. Other important liquefaction parameters are K_2 and K_7, which indicate glucose production rate and enzyme deactivation in time.

The placement of pretreatment parameters such as E_F, E_G, E_H, and E_{Ac} among the sensitive parameters shows how important pretreatment conditions are for downstream. Fermentation parameters are also numerous among the sensitive parameters. Fermentation with enhanced GMO yeast for bioethanol production is a key process in the biorefinery together with steam pretreatment. Liquefaction parameters have a lower importance because the overall hydrolysis time is long enough to compensate for any parameter uncertainties. The liquefaction process has a pure hydrolysis phase of 140 h followed by fermentation where enzymes are still active continuing cellulose degradation (simultaneous saccharification and fermentation).

The sensitivity threshold is set at 0.046, which reduces the parameters count to 22 out of 96 showing the importance of the sensitivity analysis. These parameters are then used in the following uncertainty analysis.

C.3.4 Uncertainty Analysis of Costs, Profit and Optimal Solution

The standard deviation and correlation matrix for pretreatment and enzymatic hydrolysis parameters can be obtained from [5] and [8], respectively. Regarding the

fermentation model, there is no real data that could be used to obtain parameter estimates. However, one can generate measurements through simulation, i.e. glucose, xylose and ethanol levels, add measurement noise, and then run a parameter estimation procedure for estimating the standard deviation and correlation matrix for these parameters, which is done here in this study.

Latin Hypercube Sampling (LHS) with correlation control can then be performed for all model parameters. Feedstock composition is sampled with uniform LHS considering a 5 % variation in composition. $N = 44$ samples are extracted for model and feed parameters. The Monte Carlo simulations are performed by running Algorithm 5 for each set of parameters. The simulation outputs are collected, and the 5th, 50th and 95th percentiles are then calculated for profit, costs, and optimal solution.

The uncertainty analysis is carried separately for feed, and then for model parameters in order to observe the contribution of each source of uncertainty. A last analysis combines the samples of feed and model parameters to find the overall effect of the uncertainty sources on the outputs.

The entire sensitivity and uncertainty analysis for combined model and feed parameters is summarized in Algorithm 6. To run a separate model, and then feed uncertainty analysis, the algorithm can be trivially modified either by keeping C_b or θ constant in step 11.

Algorithm 6 Sensitivity and uncertainty analysis with combined model and feed parameters.

1: Initialize model parameters θ and feed composition C_b.
2: Optimal deterministic solution: $[T_{tr}, F_e, M_y] \leftarrow$ (C.43).
3: Sensitivity analysis of the cost function in $[T_{tr}, F_e, M_y]$: calculate δ_k as in equation (C.34).
4: Rank all parameters with respect to δ_k.
5: Select a subset θ_R such that θ_R is above a threshold.
6: Set standard deviations and correlation matrices for θ_R.
7: $\theta \leftarrow$ LHS of θ_R with correlation control to generate N sets of model parameters.
8: Set bounds for feedstock composition.
9: $C_b \leftarrow$ Uniform LHS for feedstock composition to generate N sets of compositions.
10: **for** Each set of model and feed parameters **do**
11: $\left[z_i, c_{p_i}, c_{eh_i}, c_{f_i}, c_i\right] = \text{COSTS}(\theta_i, C_{b_i})$
12: Calculate the 5th, median and 95th percentile for profit, costs, and optimal solution.

The results of the uncertainty analysis are displayed in Figure C.5 and commented

below:

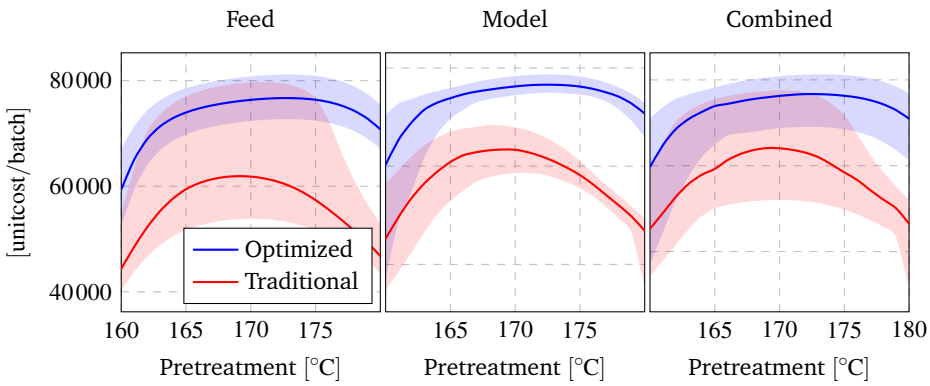
- The profit curve is drawn in Figure C.5(a), which is used to identify the optimal operational point. The traditional biorefinery operation is to follow a fixed recipe with little adjustments to pretreatment conditions. This traditional recipe is most often found by offline experiments on decoupled refinery steps that do not take into account the interactions between the conversion stages and utilities prices. This is why the traditional operation is sub-optimal from an economic point of view. In contrast, the optimization layer is capable of adapting to pretreatment temperatures and finds the optimal operation by considering the integrated process. The optimized operation is superior to a traditional recipe with a higher mean profit curve at any pretreatment temperature.

  At low temperatures, most of the uncertainty is due to model parameters, but it shifts after 165 °C when feed uncertainty becomes dominant. The traditional operation is highly affected by feed uncertainty, while the optimized operation has a reduced uncertainty on the profit curve.

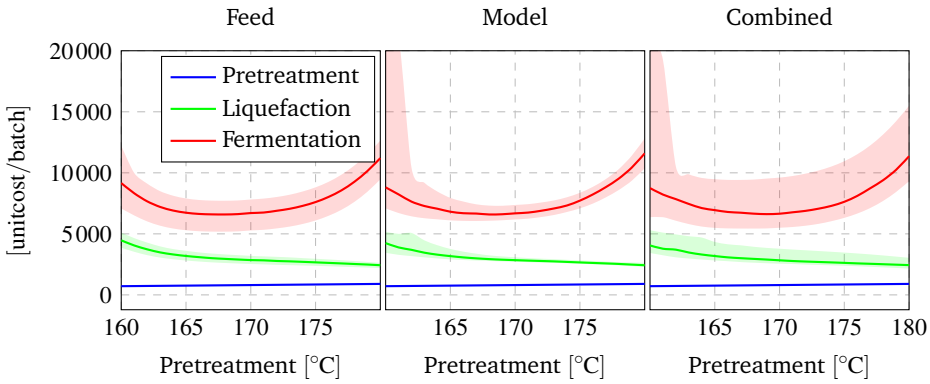
  Another important result is that the optimized profit curve is flatter than the traditional curve allowing a wider range of operation with little impact on profit value. The optimal operational point can be picked as the maximum point on the median profit curve, and lies between 171 °C to 176 °C. The optimal refinery operates at around 18 % higher profit than a traditional plant without an optimization layer.

- Figure C.5(b) shows the refinery costs split into pretreatment, liquefaction and fermentation as a function of pretreatment conditions. From left to right, the uncertainty analysis is carried with respect to separate feed and model parameters (left and center plots), and combined parameters (right plot). The pretreatment costs are only due to the steam used in the thermal reactor. The biorefinery is considered to be integrated with a local power plant, possibly owned by the same company following the IBUS principle [4]. Such a design lowers the cost of steam significantly. A higher pretreatment temperature demands more steam but the overall increase in cost for modifying the temperature from 160 °C to 180 °C is approximately 100 unitcost, which is negligible compared to enzymatic hydrolysis or fermentation.

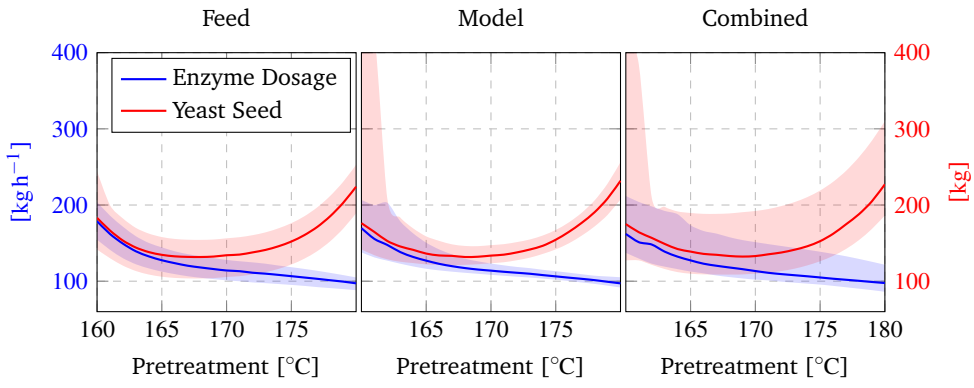
  Liquefaction costs are high at low temperature because: (1) the biomatrix is not sufficiently opened to expose the whole cellulose, and (2) there is a large amount of unhydrolyzed hemicellulose, which leads to a high production of



(a) Potential profit with uncertainty for one fermentation batch.



(b) Refinery costs for each biomass conversion step.



(c) Solution of the optimization problem.

Figure C.5: Optimal costs, potential profit, and solution of the optimization problem.

xylooligomers and xylose that inhibit the enzymatic hydrolysis further. In order to compensate for these negative effects, both the enzyme and yeast dosage are increased. The liquefaction costs decrease as the pretreatment temperature increases, which makes sense as the biomatrix opens significantly to expose cellulose, and also hemicellulose is partially removed from the enzymatic hydrolysis process.

Fermentation costs have the shape of a convex curve due to: (1) at low pretreatment temperatures a higher yeast seed could contribute to a faster digestion of sugars, which enhances the saccharification process from fermentation by reducing the C5 sugars inhibition leading to a higher ethanol yield; (2) at high temperatures the amount of inhibitors negatively affect fermentation but more yeast could compensate for the inhibitory effects of the pretreatment degradation products.

Feed uncertainty is rather constant through the entire temperature range. Uncertainty due to model parameters is high at low temperatures where the biomatrix opening highly affects the cost range. After 165 °C the model uncertainty is significantly reduced becoming lower than the feed. The combined model and feed uncertainty indicate high uncertainty at low temperatures when the pretreatment is insufficient.

- Figure C.5(c) illustrates the optimal solution as a function of pretreatment temperatures. Enzyme dosage is expressed in $\mathrm{kg\,h^{-1}}$, while yeast seed is given in kg. Uncertainty is higher at lower temperatures when the biomass is not sufficiently pretreated, and remains relatively constant once the biomatrix opens. Also, feed uncertainty has a higher impact than model parameters after 165 °C.

  Increasing pretreatment temperatures is beneficial for enzymatic hydrolysis as it lowers the necessary enzyme dosage, but is negative for fermentation as the amount of inhibitors rises with temperature. Also, around the optimal operational point, uncertainty due to feed parameters dominates that of model parameters.

In reality there are several factors that can degrade the performance of the optimization layer, and should be accounted for in real implementation. The feed rate in this study case was set to a low value, which does not allow inhibitors accumulation in the fermentation tank. However, at higher feed rates inhibitors accumulation becomes a bottleneck, which can be counteracted by calculating an optimal feed rate profile [86]. This study also disregards the temperature

dependence of the yeast performance. In reality the enzymatic hydrolysis and fermentation processes run at different optimal temperatures. The solution is to calculate a temperature profile for finding the best trade-off between the two processes [79, 85].

C.3.5 Stochastic Optimization Solution

Finding the optimal point by running the process through a wide range of pretreatment temperatures requires a long computational time, and is not feasible in an industrial application. A better way is to embed the feed and model parameters uncertainty into the objective function, and pick the mean cost value:

$$\begin{aligned} \max_{T_{tr},F_e,M_y} \quad & \tfrac{1}{N}\sum_{1}^{N}\left[M_{Eth}(t_f)P_{Eth} - (F_sP_s + F_eP_e + M_yP_y)\right] \\ \text{subject to} \quad & 0 = f(x(t),u(t)) \\ & \dot{x}_f = h(x_f,u_f) \\ 150 \;&\leq\; T_{tr} \;\leq\; 210\,^{\circ}\mathrm{C} \\ 10 \;&\leq\; F_e \;\leq\; 1000\,\mathrm{kg\,h^{-1}} \\ 10 \;&\leq\; M_y \;\leq\; 1000\,\mathrm{kg} \end{aligned} \tag{C.47}$$

N is the number of parameter sets generated through LHS with correlation control, or the Monte Carlo simulations count. The cost function is evaluated by running N simulations selecting the profit mean value. The following optimal solution and cost value are found:

$$z_{so} = \begin{bmatrix} T_{tr} \\ F_e \\ M_y \end{bmatrix} = \begin{bmatrix} 171.5\,^{\circ}\mathrm{C} \\ 113\,\mathrm{kg\,h^{-1}} \\ 146\,\mathrm{kg} \end{bmatrix} \qquad c_{so} = 7.6015 \times 10^4\,\text{unitprofit} \tag{C.48}$$

where z_{so} is the optimal solution in the stochastic optimization case, and c_{so} is the value of the cost function. Solution (C.48) is relatively close to the one found in the deterministic case from (C.44) with a slightly lower profit value.

C.3.6 Deterministic Simulations at the Optimal Point

A deterministic simulation is run corresponding to the optimal operational point from solution (C.44). The pretreatment and the enzymatic hydrolysis are continuous processes and the steady state values at the optimal point are shown in Table C.5. The pretreated fibers are rich in cellulose and have a dry matter content of about 35 % as suggested by [55] for an efficient liquefaction process. Most solubles were separated from the fibers in the pretreatment process before liquefaction. The remaining hemicellulose continues to be degraded to sugars in the enzymatic hydrolysis tanks.

When the level of C5 sugars increase, they strongly inhibit glucose production and a part of cellulosic fibers remain in solid state. This is why the liquefied fibers still contain cellulose before fermentation, i.e. $50\,g\,kg^{-1}$, approximately 30 % of the initial cellulose content. The remaining cellulose continues conversion to glucose in the fermentation tank where enzymes are still active.

Table C.5: Steady states for pretreatment and enzymatic hydrolysis at the optimal point.

	Pretreated fibers	C5 liquid	Liquefied fibers	Fermentation
Flow/Mass	$2316\,kg\,h^{-1}$	$628\,kg\,h^{-1}$	$2487\,kg\,h^{-1}$	220 t
Composition	$1000\,g\,kg^{-1}$	$1000\,g\,kg^{-1}$	$1000\,g\,kg^{-1}$	$1000\,g\,kg^{-1}$
Cellulose	146	1.2	50	4.4
Xylan	60	0.5	1	0
Arabinan	0	0	0	0
Lignin	85	0.7	78	60
Acetyls	16	0.1	0.1	0
Ash	6	18	5.7	7.8
Acids	1.5	4.1	16	0
Glucose	3.5	10	98	0
Xylooligomers	0.5	1.2	5.8	0.1
Xylose	10	29.7	59	0
Arabinose	5	15.5	5	0
Furfural	0.2	0.5	0.2	0
5-HMF	0.1	0.3	0.1	0
Base	0	0	6.6	9.5
Enzymes	0	0	4.9	2.4
Biomass	0	0	0	8.4
Ethanol	0	0	0	79
CO_2	0	0	0	80
Water	645	918	643	702
Other	21.2	0.2	26.6	46.4
Temperature	50 °C	50 °C	50 °C	35 °C

Figure C.6 shows the fermentation batch process at the optimal point. The top plot illustrates C6 and C5 sugars depletion, ethanol production, and biomass growth. The bottom plot displays remaining cellulose and xylan conversion during simultaneous saccharification and fermentation. In the inoculum phase (first 10 h) the yeast concentration is high but as the fed-batch phase starts, biomass concentration is diluted in liquefied fibers from the enzymatic hydrolysis and C5 liquid from the pretreatment process. Ethanol production has several stages: (1) formation on

glucose consumption till around 100 h, (2) production based on xylose consumption till 170 h, (3) as xylose is depleted, its inhibition on enzymatic hydrolysis disappears and glucose production from simultaneous saccharification and fermentation is accelerated in the last 20 h.

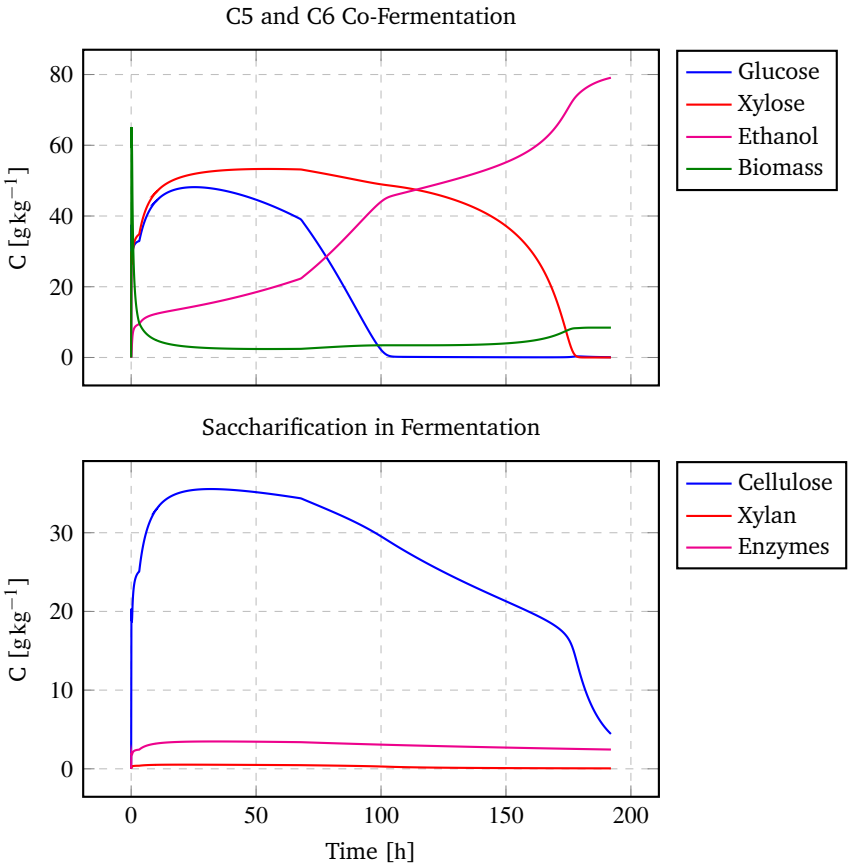


Figure C.6: Simultaneous saccharification and C5-C6 co-fermentation.

C.4 Conclusions

This was a study on economical optimization of a large scale second generation biorefinery in a simulated environment. The proposed optimization procedure is based on steady-state models (pretreatment and enzymatic hydrolysis), and dynamic fermentation model. Optimal operation depends on pretreatment temperature, enzyme dosage in liquefaction, and yeast seed in fermentation. Increasing pretreatment temperature is positive for the performance of the enzymatic hydrolysis while

negative for ethanol yield. Uncertainties in kinetics of pretreatment, liquefaction and fermentation were negligible on the economic objective function around the optimal operational point. The main source of uncertainty was found to be the inflow feed composition. The optimization layer reduced the uncertainty and flattened the profit curve allowing a wider range of operation with higher profit. The overall improvement of the optimization layer is of approximately 18 % over the traditional plant.

Supplementary Material

Introduction

This supplementary material summarizes the dynamic mathematical models for pretreatment and enzymatic hydrolysis used in the plantwide optimization study. These models are necessary for calculating the steady state values. Due to the high complexity and model nonlinearities, an analytical steady state solution is hard to find. The alternative is to run a sufficiently long simulation with constant inputs and grab the final stabilized state values.

Mathematical Models

Pretreatment

The pretreatment process occurs in a large horizontal thermal reactor, which is pressurized with steam till the necessary reaction pressure/temperature. The following parameter distributed equation models the biomass transportation from inlet to outlet subject to conversion due to the steam temperature [5]:

$$\frac{dC_k}{dt} = \frac{u_z}{\delta_z}(C_{k-1} - C_k) + R_k \tag{C.49}$$

C_k is the composition vector in central cell k, C_{k-1} is the composition vector from the western neighbor, and R_k is the reaction rate vector from current cell k. Movement from west to east (left to right) is assumed.

The pretreatment conversion mechanism is shown in Figure C.7. The reaction rates for building vector R_k are modeled as first order Arrhenius type reactions, and are detailed in Prunescu et al. [5]. A summary is given below:

- Glucose production due to cellulose hydrolysis:

$$r_G = k_G \exp\left(-\frac{E_G}{R \cdot T_K}\right) C_{C_S} \tag{C.50}$$

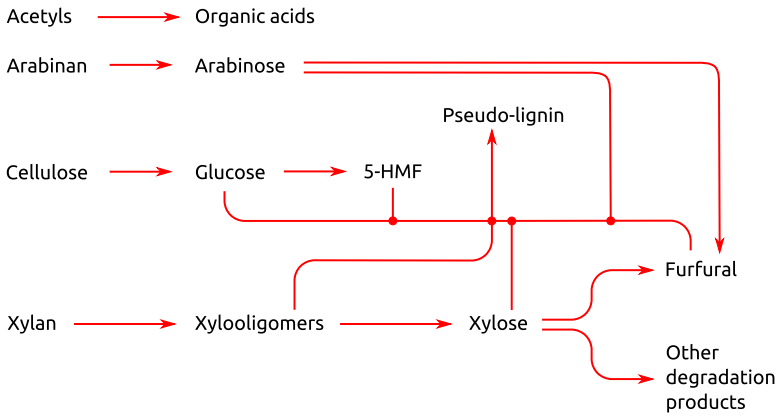


Figure C.7: Hydrothermal pretreatment conversion mechanism [5].

where k_G is th reaction rate constant, E_G is the activation energy, and C_{C_S} is the cellulose is solid state. R and T_K are the universal gas constant, and the cell temperature in Kelvin degrees.

- 5-HMF production due to glucose degradation:

$$r_H = k_H \exp\left(-\frac{E_H}{R \cdot T_K}\right) C_G \tag{C.51}$$

where C_G is the glucose concentration.

- Arabinan C_{A_S} is in low concentration and fully hydrolyzes to arabinose following the rate:

$$r_A = k_A \exp\left(-\frac{E_A}{R \cdot T_K}\right) C_{A_S} \tag{C.52}$$

- Xylan is hydrolyzed having xylooligomers as an intermediate product:

$$r_{Xo} = k_{Xo} \exp\left(-\frac{E_{Xo}}{R \cdot T_K}\right) C_{X_S} \tag{C.53}$$

where C_{X_S} is xylan in solid state.

- Xylooligomers are further decomposed into xylose:

$$r_X = k_X \exp\left(-\frac{E_X}{R \cdot T_K}\right) C_{Xo} \tag{C.54}$$

with C_{Xo} as the concentration of xylooligomers.

- Xylose C_X and arabinose C_A are C5 sugars and can degrade into furfural:

$$r_F = k_F \exp\left(-\frac{E_F}{R \cdot T_K}\right) (C_X + C_A) \tag{C.55}$$

- Carbohydrates can react with inhibitors (furfural and 5-HMF) to create spherical droplets called pseudo-lignin [53]:

$$r_L = k_L \exp\left(-\frac{E_L}{R \cdot T_K}\right)(C_{Xo} + C_X + C_A + C_G)(C_F + C_H) \tag{C.56a}$$

$$r_L = r_{L_{Xo}} + r_{L_X} + r_{L_A} + r_{L_G} \tag{C.56b}$$

$$r_L = r_{L_F} + r_{L_H} \tag{C.56c}$$

$r_{L_{Xo}}$, r_{L_X}, r_{L_A}, and r_{L_G} show pseudo-lignin production due to xylooligomers, xylose, arabinose, and glucose participation expressed separately [5]:

$$r_{L_{Xo}} = k_L \exp\left(-\frac{E_L}{R \cdot T_K}\right) C_{Xo}(C_F + C_H) \tag{C.57a}$$

$$r_{L_X} = k_L \exp\left(-\frac{E_L}{R \cdot T_K}\right) C_X(C_F + C_H) \tag{C.57b}$$

$$r_{L_A} = k_L \exp\left(-\frac{E_L}{R \cdot T_K}\right) C_A(C_F + C_H) \tag{C.57c}$$

$$r_{L_G} = k_L \exp\left(-\frac{E_L}{R \cdot T_K}\right) C_G(C_F + C_H) \tag{C.57d}$$

r_{L_F} and r_{L_H} express pseudo-lignin production with furfural and 5-HMF participation [5]:

$$r_{L_F} = k_L \exp\left(-\frac{E_L}{R \cdot T_K}\right)(C_{Xo} + C_X + C_A + C_G)C_F \tag{C.58a}$$

$$r_{L_H} = k_L \exp\left(-\frac{E_L}{R \cdot T_K}\right)(C_{Xo} + C_X + C_A + C_G)C_H \tag{C.58b}$$

- Hemicellulose contains acetyls C_{Ac_S} that hydrolyze to produce acetic acid:

$$r_{Ac} = k_{Ac} \exp\left(-\frac{E_{Ac}}{R \cdot T_K}\right) C_{Ac_S} \tag{C.59}$$

The composition vector C_k and the reaction rate array R_k from Equation (C.49)

then become:

$$C_k = \begin{bmatrix} C_{C_S} \\ C_{X_S} \\ C_{A_S} \\ C_{L_S} \\ C_{Ac_S} \\ C_G \\ C_{Xo} \\ C_X \\ C_A \\ C_{Ac} \\ C_F \\ C_H \\ C_W \\ C_O \end{bmatrix} \quad R_k = \begin{bmatrix} -r_G \\ -r_{Xo} \\ -r_A \\ r_L \\ -r_{Ac} \\ r_G - r_{O_G} - (1-\alpha) r_{L_G} \\ r_{Xo} - r_X - (1-\alpha) r_{L_{Xo}} \\ r_X - r_{F_X} - r_{O_X} - (1-\alpha) r_{L_X} \\ r_A - r_{O_A} - r_{F_A} - (1-\alpha) r_{L_A} \\ r_{Ac} \\ r_F - \alpha r_{L_F} \\ r_H - \alpha r_{L_H} \\ 0 \\ r_{O_X} + r_{O_G} + r_{O_A} \end{bmatrix} \tag{C.60}$$

The mass balance is ensured by the following conditions:

$$\sum R_k = 0\,\mathrm{g/(kgs)} \qquad \sum C_k = 1000\,\mathrm{g\,kg^{-1}} \tag{C.61}$$

Enzymatic Hydrolysis

The enzymatic hydrolysis model has been formulated and analyzed in Prunescu and Sin [6]. The competitive conversion routes are presented in Figure C.8. The model has been extended in this study with enzyme deactivation in time as suggested by Zhang et al. [61] plus enzymatic efficiency with respect to biomatrix opening described by the severity factor. A summary of the model equations is given next:

- The mass balance is built similarly to the thermal reactor case because the enzymatic hydrolysis runs at a very high initial dry matter, i.e. 40 %. If Equation (C.49) is discretized into a single cell then it is equivalent to a continuous stirred tank reactor (CSTR). The enzymatic hydrolysis occurs in several tanks interconnected in series. In the first tank viscosity has a significant drop and is discretized in $N = 6$ cells, while the other tanks behave as CSTRs such that the retention time meets the 140 h constraint.

- The enzymatic solution parametrization:

$$C_E = C_{E_C} + C_{E_X} \tag{C.62}$$

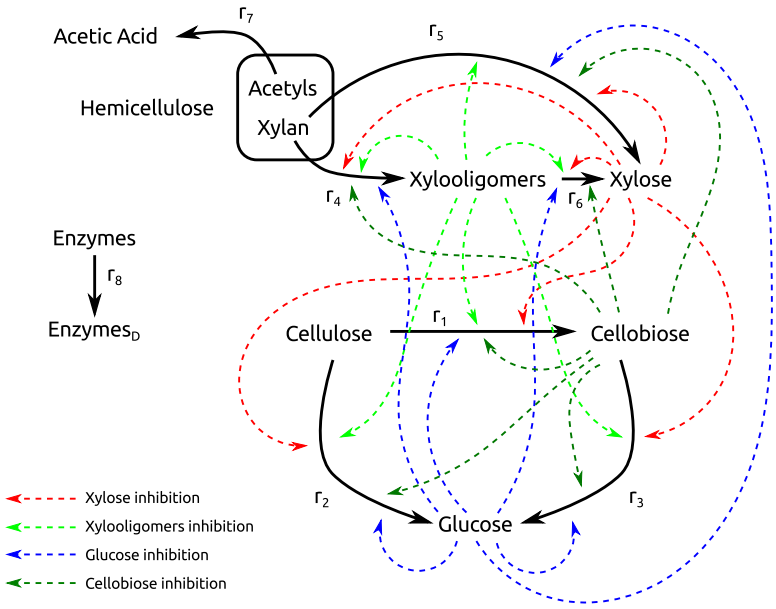


Figure C.8: Enzymatic hydrolysis conversion mechanism with inhibition. Extended from Prunescu and Sin [6] with xylooligomers intermediate product, acetyl groups, and enzyme deactivation.

The enzymatic solution contains cellulase C_{E_C} and xylanase C_{E_X}. Cellulase is made up of endo-exo type cellulase $C^E_{E_C}$ and β-glucosidase $C^G_{E_C}$:

$$C_{E_C} = C^E_{E_C} + C^G_{E_C} \tag{C.63}$$

Xylanase consists of exo-endo type xylanase $C^E_{E_X}$ and xylosidase $C^X_{E_X}$:

$$C_{E_X} = C^E_{E_X} + C^X_{E_X} \tag{C.64}$$

The total concentration of enzymes can be parametrized using fractions:

$$C_E = \alpha^E_C C_E + \alpha^G_C C_E + \alpha^E_X C_E + \alpha^X_X C_E \tag{C.65}$$

where α^E_C, α^G_C, α^E_X, and α^X_X are fractions of each type of cellulase and xylanase. Enzymes can be in two states: bounded or free. There is an equilibrium between the states described by:

$$C^E_{E_{C_B}} = E^E_{M_C} \frac{K^E_{A_C} C^E_{E_{C_F}}}{1 + K^E_{A_C} C^E_{E_{C_F}}} C_S \tag{C.66}$$

where $C^E_{E_{C_B}}$ are bounded exo-endo type cellulase, $C^E_{E_{C_F}}$ are free same type enzymes. $E^E_{M_C}$ is a maximum adsorption term, and K_{A_C} the Langmuir adsorption

constant. C_S is the solid substrate that enzymes are bound to. Equation (C.66) applies to all types of enzymes, i.e. β-glucosidase, endo-exo type xylanase, and xylosidase.

- Kinetic modeling of the reaction rates from Figure C.8: Reaction r_1 describes cellulose hydrolysis to cellobiose:

$$r_1 = \frac{K_1 \eta(T,pH,r) C^E_{E_{C_B}} C_{C_S}}{1 + \frac{C_C}{I_{C_1}} + \frac{C_X}{I_{X_1}} + \frac{C_G}{I_{G_1}} + \frac{C_{Xo}}{I_{Xo_1}} + \frac{C_{Eth}}{I_{Eth_1}}} \tag{C.67}$$

K_1 is the reaction rate, $\eta(T,pH,r)$ is the temperature, pH, and severity dependency of the enzymatic activity, $C^E_{E_{C_B}}$ are the bounded exo-endo type cellulase, and C_{C_S} is the concentration of solid cellulose. The reaction rate is inhibited by cellobiose C_C, xylose C_X, xylooligomers C_{Xo}, and glucose C_G through inhibition terms I_{C_1}, I_{X_1}, I_{Xo_1}, and I_{G_1}. In simultaneous saccharification and fermentation (SSF) ethanol also inhibits cellobiose formation [96] modeled through I_{Eth_1} in the above rate.

Cellulose is also decomposed straight to glucose by the action of both endo-exo type enzymes and β-glucosidase:

$$r_2 = \frac{K_2 \eta(T,pH,r) \left(C^E_{E_{C_B}} + C^G_{E_{C_B}} \right) C_{C_S}}{1 + \frac{C_C}{I_{C_2}} + \frac{C_{Xo}}{I_{Xo_2}} + \frac{C_X}{I_{X_2}} + \frac{C_G}{I_{G_2}} + \frac{C_{Eth}}{I_{Eth_2}}} \tag{C.68}$$

Cellobiose decomposes to glucose:

$$r_3 = \frac{K_3 \eta(T,pH) C^G_{E_{C_F}} C_C}{I_3 \left(1 + \frac{C_{Xo}}{I_{Xo_3}} + \frac{C_X}{I_{X_3}} + \frac{C_G}{I_{G_3}} + \frac{C_{Eth}}{I_{Eth_3}} \right) + C_C} \tag{C.69}$$

Xylan hydrolysis follows a similar path. Xylan form a xylooligomers pool:

$$r_4 = \frac{K_4 \eta(T,pH) C^E_{E_{X_B}} C_{X_S}}{1 + \frac{C_C}{I_{C_4}} + \frac{C_{Xo}}{I_{Xo_4}} + \frac{C_X}{I_{X_4}} + \frac{C_G}{I_{G_4}} + \frac{C_{Eth}}{I_{Eth_4}}} \tag{C.70}$$

Xylan could also be decomposed straight to xylose in a smaller amount:

$$r_5 = \frac{K_5 \eta(T,pH) \left(C^E_{E_{X_B}} + C^X_{E_{X_B}} \right) C_{X_S}}{1 + \frac{C_C}{I_{C_5}} + \frac{C_{Xo}}{I_{Xo_5}} + \frac{C_X}{I_{X_5}} + \frac{C_G}{I_{G_5}} + \frac{C_{Eth}}{I_{Eth_5}}} \tag{C.71}$$

Xylooligomers are further decomposed to xylose:

$$r_6 = \frac{K_6 \eta(T,pH) C^X_{E_{X_F}} C_{Xo}}{I_6 \left(1 + \frac{C_C}{I_{C_6}} + \frac{C_X}{I_{X_6}} + \frac{C_G}{I_{G_6}} + \frac{C_{Eth}}{I_{Eth_6}} \right) + C_{Xo}} \tag{C.72}$$

Acetic acid production happens due to acetyls being released along with xylan hydrolysis:

$$r_7 = \beta_{Ac}(r_4 + r_5) \tag{C.73}$$

Enzymes deactivate in time due to thermal inactivation and exposure to ethanol [61]:

$$r_8 = -K_8 C_E^2 \tag{C.74}$$

- pH is modeled based on the charge balance equation:

$$[\mathrm{H}^+] - [\mathrm{OH}^-] - [\mathrm{Ac}^-] + [\mathrm{Na}^+] = 0 \tag{C.75}$$

The hydrogen ion concentration is found as the solution to the charge balance equation. The other ions are expressed using the states and dissociation constants. pH is then defined as:

$$pH = -\log_{10}[\mathrm{H}^+] \tag{C.76}$$

- The enzymatic activity is a function of temperature, pH and severity factor as illustrated in Figure C.9:

$$\eta(T, pH, r) = \eta_T(T)\eta_p(pH)\eta_r(r) \tag{C.77}$$

The temperature and pH dependency are determined based on linear interpolation of experimental tabular data, and have the shape of a bell with a single optimal peak [57, 58]. The severity [50, 97] dependency illustrate the biomatrix opening, and models how hard the enzymes can access the cellulosic and hemicellulosic fibers due to the physical structure of the biomass. The severity factor is a function of retention time and temperature [97]. Insufficient pretreatment can block enzymes accessibility to fibers.

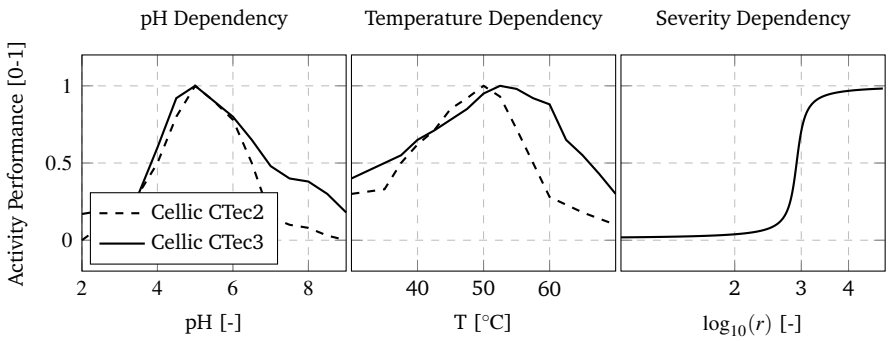


Figure C.9: Enzymatic activity dependency on temperature, pH, and severity factor r.

The composition vector C_k and the reaction rates array R_k are shown next:

$$C_k = \begin{bmatrix} C_{C_S} \\ C_{X_S} \\ C_{L_S} \\ C_{Ac_S} \\ C_{Ac} \\ C_C \\ C_G \\ C_{Xo} \\ C_X \\ C_F \\ C_H \\ C_B \\ C_E \\ C_W \\ C_O \end{bmatrix} \qquad R_k = \begin{bmatrix} -r_1 - r_2 \\ -r_4 - r_5 \\ 0 \\ -r_7 \\ r_1 - r_3 \\ r_2 + r_3 \\ r_4 - r_6 \\ r_5 + r_6 \\ r_7 \\ 0 \\ 0 \\ 0 \\ -r_8 \\ 0 \\ r_8 \end{bmatrix} \tag{C.78}$$

Conservation of mass is ensured by the following conditions:

$$\sum R_k = 0\,\mathrm{g/(kgs)} \qquad \sum C_k = 1000\,\mathrm{g\,kg^{-1}} \tag{C.79}$$

Paper D

Modeling and L1 Adaptive Control of Temperature in Biomass Pretreatment

Remus Mihail Prunescu[1], Mogens Blanke[*1], Gürkan Sin[2]

[1]Department of Electrical Engineering, Automation and Control Group, Technical University of Denmark, Elektrovej Building 326, 2800, Kgs. Lyngby, Denmark

[2]CAPEC-PROCESS, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads Buildings 227 and 229, 2800, Kgs. Lyngby, Denmark

Abstract:
Biomass steam pretreatment is a key process in converting agricultural wastes to bioethanol. The pretreatment occurs in a large pressurized tank called a thermal reactor. Two key parameters influence the successfulness of the process: the reactor temperature, and the retention time. A particle pump pressurizes untreated biomass from atmospheric to reactor pressure with recycled steam from the reactor. This paper formulates a steam mathematical model both for the thermal reactor and the particle pump, which is then used to design an L1 adaptive output feedback controller for the reactor temperature. As steam is recycled from the reactor to pressurize the particle pump, pressure drops and the reactor temperature is disturbed. The main control challenge is to reject these disturbances and keep a steady temperature. The nonlinear process model embeds mass and energy balances, valve characteristics, and enthalpy-pressure and pressure-temperature dependencies. Nonlinear feed-forward terms are added in the control strategy. The process model, the control strategy, the application of the L1 adaptive controller and its tuning method based on minimizing a cost function represent novelties of this paper.

[*]Principal corresponding author. Tel.: +45 45253565; E-mail: mb@elektro.dtu.dk

D.1 Introduction

Biofuel receives more interest lately due to the increase in oil price worldwide, and due to the green commitments that governments have taken for reducing emissions of greenhouse gases, e.g. the Kyoto protocol [19]. In this context, many private companies and governments invest in large scale biofuel production plants [3].

The conversion of agricultural wastes to ethanol requires several steps. A thermal reactor pretreats the biomass by removing the wooden part of the plant, i.e. lignin, and creates a mixture rich in cellulose fibers. Enzymes breakdown the fibers in a liquefaction process and large tanks store the resulting slurry for fermentation. Distillation columns recover ethanol, which is used in preparation of fuel blends. Two more by-products are obtained: lignin, which is recovered from the pretreatment process as bio-pallets and co-combusted in a power plant; and C5 molasses, which is a syrup rich in nutrients and sold to farmers for feeding their animals. The entire biofuel production cycle is thoroughly described in [3, 4].

The biomass pretreatment process is the key step in biofuel production. There are various methods of pretreatment, most of them involving steam, strong acids or weak acids [121]. The effects of different pretreatment methods were investigated in [44, 122]. A combination of steam and weak acids gives the best results. Also, to reduce steam operating costs, refineries are integrated with a power plant following the IBUS principle [4].

The steam pretreatment process occurs in a pressurized continuous thermal reactor, which is preceded by a pressurization unit also known as a particle pump [71]. Depending on the load, the particle pump releases an amount of biomass to the thermal reactor with a certain frequency. The degree of pretreatment determines the chemical composition of the outstream. Steam pretreatment started to be investigated in [97] and an empirical pretreatment model for ethanol production was formulated in [50]. Two key parameters of the process, i.e. retention time and reactor temperature, are vital to an effective process.

Various disturbances have been identified for a large thermal reactor in [71], among which the most important one is the particle pump operation cycle, which is causing oscillations in the steam layer temperature. These oscillations further disturb the temperature of the biomass layer, causing an irregular pretreatment process. Temperature matters because a deviation of $10\,^{\circ}\mathrm{C}$ can cause a drop of $10\,\%$ in cellulose recovery [50]. A poor pretreatment can also lead to formation of inhibitors that affect the downstream processes of enzymatic hydrolysis and fermentation, or it could disturb the pH level of the outstream by releasing more

or less acetic acid from the biomass. It is well known that enzymes are sensitive to pH following a bell-shaped curve [123]. Therefore, the main control challenge in this investigation is to reject any disturbances that affect the reactor temperature in order to ensure a uniform temperature environment in the steam and biomass layers.

To achieve this objective, an adaptive control strategy for the pretreatment temperature is proposed in this article. First, a nonlinear model of the steam layer is formulated from mass and energy balances, enthalpy-pressure and enthalpy-temperature dependencies, and nonlinear valve characteristics. Disturbance sources are also introduced, i.e. condensation of steam and uncertainties in the measurement of recycled steam. The biomass load in the reactor is also a factor and the controller will be tuned for a specific load and tested in another operational point. A disturbance can take the system to multiple operational points and adaptation should improve the control performances. Therefore, an adaptive control strategy is developed based on the L1 adaptive output feedback controller. L1 adaptive control represents the latest novelty in control theory [10]. Also, a new tuning method of the L1 controller is proposed in this paper based on minimizing the integral absolute error (IAE) performance function. The modelling work, the application of the L1 adaptive controller on a biomass pretreatment process and the tuning method of the controller have not been reported earlier in the literature.

D.2 Process Description

The process has been described in [71] and is briefly reiterated in this section. A process diagram is illustrated in Figure D.1. The particle pump receives small amounts of soaked biomass and pressurizes them till the reactor pressure. Afterward, the particle pump releases the biomass into the thermal reactor and the cycle repeats. In the thermal reactor, the biomass is pushed horizontally by a snail. The particle pump operates at various frequencies depending on the desired load of the thermal reactor. A fast operation increases the load in the reactor assuming the snail speed constant. When the load increases the available volume for the steam layer diminishes. A reduced air volume can be pressurized faster by steam, so the time constants of the system change according to the load.

It is assumed that the particle pump is pressurized only with recycled steam from the reactor while the thermal reactor is pressurized only with fresh steam, which enters the tank through the bottom. It is preferred to use recycled steam when pressurizing the particle pump because it would create a pressure disturbance that causes a burst of steam from the bottom, which would facilitate the breakdown of

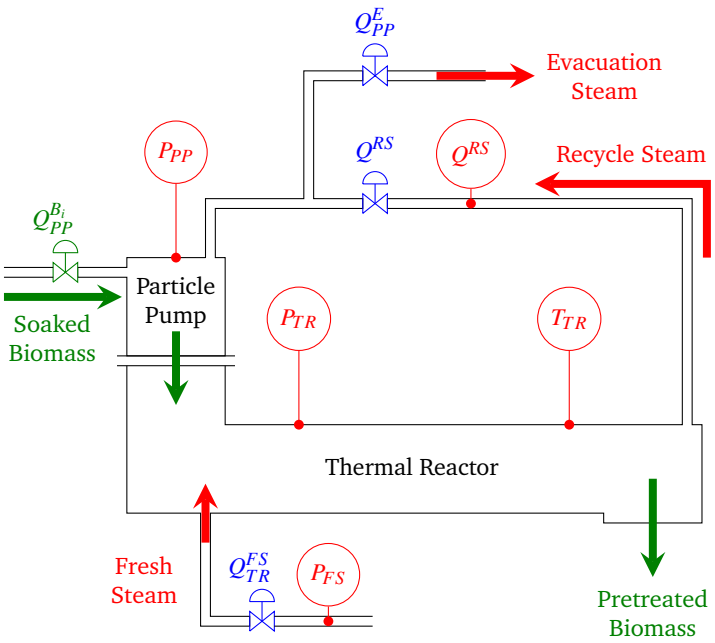


Figure D.1: Process diagram with instrumentation. Green arrows follow the biomass path while red arrows indicate steam flow. There are 3 steam valves, i.e. Q^{FS} or fresh steam valve, Q^{RS} or recycled steam valve and Q^E or evacuation valve. P_{PP}, P_{TR} and P_{FS} measure the pressure in the particle pump, thermal reactor and fresh steam pipe, respectively. T_{TR} indicates the reactor temperature.

soaked biomass leading to a more uniform pretreatment process.

The pressure in the particle pump and the pressure in the reactor are measured as P_{PP} and P_{TR}. Another pressure sensor is positioned in the fresh steam pipe and the measurement is denoted as P_{FS}. The fresh steam arrives from a flash tank and is assumed to be saturated. The temperature in the thermal reactor is of interest and directly measured as T_{TR}. The recycle steam flow is also measured as Q^{RS} and is necessary to construct a feed-forward action. The steam layers from the particle pump and thermal reactor are assumed to be uniform. There are 3 steam valves that can be manipulated and their strokes are denoted as S^E_{PP}, S^{FS}_{TR} and S^{RS}.

D.3 Mathematical Model

Table D.1: Process model nomenclature.

States		
m_{PP}^{S}	Steam mass in particle pump (PP)	kg
h_{PP}^{S}	Specific steam enthalpy in PP	$\mathrm{kJ\,kg^{-1}}$
m_{TR}^{S}	Steam mass in thermal reactor (TR)	kg
h_{TR}^{S}	Specific steam enthalpy in TR	$\mathrm{kJ\,kg^{-1}}$
m_{PP}^{B}	Mass of biomass in PP	kg
m_{TR}^{B}	Mass of biomass in TR	kg
Inputs		
S_{TR}^{FS}	Stroke of fresh steam valve in TR	%
S^{RS}	Stroke of recycle steam valve	%
S_{PP}^{E}	Stroke of evacuation valve in PP	%
$Q_{PP}^{B_i}$	Inflow of biomass in PP	$\mathrm{kg\,s^{-1}}$
$Q_{TR}^{B_i}$	Inflow of biomass in TR	$\mathrm{kg\,s^{-1}}$
$Q_{TR}^{B_o}$	Outflow of biomass from TR	$\mathrm{kg\,s^{-1}}$
h_{FS}^{S}	Fresh steam enthalpy	$\mathrm{kJ\,kg^{-1}}$
P_{FS}	Fresh steam pressure	bar
Outputs		
P_{PP}	Pressure in PP	bar
P_{TR}	Pressure in TR	bar
P_A	Atmospheric pressure	bar
T_{PP}	Temperature in PP	°C
T_{TR}	Temperature in TR	°C
Auxiliary Variables		
Q^{RS}	Mass flow of recycled steam	$\mathrm{kg\,s^{-1}}$
Q_{PP}^{E}	Mass outflow of steam from PP	$\mathrm{kg\,s^{-1}}$
Q_{TR}^{FS}	Mass inflow of fresh steam in TR	$\mathrm{kg\,s^{-1}}$
ρ_{PP}^{S}	Density of steam in PP	$\mathrm{kg\,m^{-3}}$
ρ_{TR}^{S}	Density of steam in TR	$\mathrm{kg\,m^{-3}}$
V_{PP}^{S}	Steam volume in PP	$\mathrm{m^3}$
V_{TR}^{S}	Steam volume in TR	$\mathrm{m^3}$
Constant Parameters		
V_{PP}	Total volume of PP	$\mathrm{m^3}$

V_{TR}	Total volume of TR	m^3
Disturbances		
Q^C_{TR}	Condensing steam in TR	$\mathrm{kg\,s^{-1}}$
Q^M_{RS}	Recycle steam measurement noise	$\mathrm{kg\,s^{-1}}$

D.3.1 Process Model

The process model consists of mass and energy balances of the steam layer for both the particle pump and the thermal reactor:

$$\frac{d(m^S_{PP})}{dt} = Q^{RS} - Q^E_{PP} \tag{D.1a}$$

$$\frac{d(m^S_{PP}h^S_{PP})}{dt} = Q^{RS}h^S_{TR} - Q^E_{PP}h^S_{PP} \tag{D.1b}$$

$$\frac{d(m^S_{TR})}{dt} = Q^{FS}_{TR} - Q^{RS} - Q^C_{TR} \tag{D.1c}$$

$$\frac{d(m^S_{TR}h^S_{TR})}{dt} = Q^{FS}_{TR}h^S_{FS} - (Q^{RS} + Q^C_{TR})h^S_{TR} \tag{D.1d}$$

Equation (D.1a) is the particle pump steam mass balance and Equation (D.1b) is the energy balance of the steam layer from the particle pump. All notations are explained in Table D.1. Similarly, Equation (D.1c) shows the mass balance from the thermal reactor and Equation (D.1d) contains the energy balance of the steam layer in the reactor.

The mass of soaked biomass in both the particle pump and thermal reactor are also tracked as it affects the available volume for steam expansion:

$$\frac{d(m^B_{PP})}{dt} = Q^{B_i}_{PP} - Q^{B_i}_{TR} \tag{D.2a}$$

$$\frac{d(m^B_{TR})}{dt} = Q^{B_i}_{TR} - Q^{B_o}_{TR} \tag{D.2b}$$

D.3.2 Valve Modelling

The valve characteristics can be retrieved from the valve manufacturer and shows K_V as a function of the valve opening or stroke S. K_V represents the flow of steam in $\mathrm{m^3\,s^{-1}}$ when the pressure drop across the valve is 1 bar. A typical characteristic is displayed in Figure D.2 and can be accurately approximated by a cubic polynomial:

$$K_V(S) = k_3S^3 + k_2S^2 + k_1S + k_0 \tag{D.3}$$

where k_i are the polynomial coefficients and S is the valve stroke.

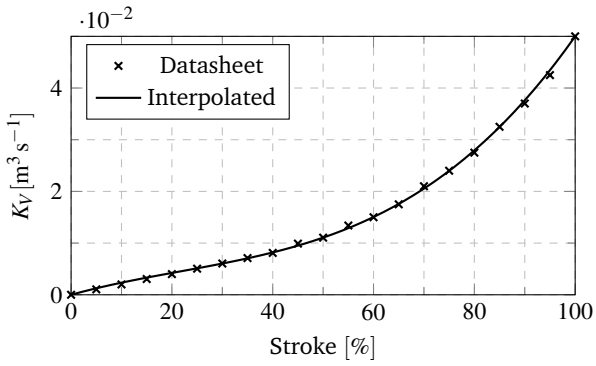


Figure D.2: Steam valve characteristics. Solid line shows the polynomial fit while bullet points indicate measurements taken from the valve datasheet.

The fresh steam is, in fact, saturated steam because it is supplied by a flash tank. The steam flow is approximated by:

$$Q^{SS} = f(S, p_i, p_o) = \begin{cases} K_V(S) \cdot 22.4 \cdot \sqrt{(p_i - p_o)p_o} & , p_o > \frac{p_i}{2} \\ K_V(S) \cdot 11.2 \cdot p_i & , p_o < \frac{p_i}{2} \end{cases} \tag{D.4}$$

where Q^{SS} is the flow of saturated steam, $K_V(S)$ is the valve characteristic given in Equation (D.3), p_i is the valve inlet pressure, and p_o is the outlet pressure expressed in bar. The recycled steam extracted from the thermal reactor is assumed to be near saturation with a flow approximated by Equation (D.4).

D.3.3 State Space Model

State vector x contains:

$$x = \begin{bmatrix} m^S_{PP} & h^S_{PP} & m^S_{TR} & h^S_{TR} & m^B_{PP} & m^B_{TR} \end{bmatrix}^{\mathrm{T}} \tag{D.5}$$

Vector u gathers all manipulated variables, i.e. the valve strokes:

$$u = \begin{bmatrix} S^{RS} & S^E_{PP} & S^{FS}_{TR} \end{bmatrix}^{\mathrm{T}} \tag{D.6}$$

The biomass flows are not considered manipulated variables for steam regulation and are placed in vector $\tilde{u}$:

$$\tilde{u} = \begin{bmatrix} Q^{B_i}_{PP} & Q^{B_o}_{PP} & Q^{B_o}_{TR} \end{bmatrix}^{\mathrm{T}} \tag{D.7}$$

Disturbances enumerates condensation effects and measurement errors of the recycle steam. Condensation occurs inside the reactor either due to heat losses to

the environment or because of temperature differences between the biomass and the steam layer. Measurement errors in the recycle steam flow may appear due to its fast and turbulent dynamics. Vector d comprises all disturbances:

$$d = \begin{bmatrix} Q_{TR}^{C} & Q_{RS}^{M} \end{bmatrix}^{\mathrm{T}} \tag{D.8}$$

The steam volume in each container is found by subtracting the biomass volume from the total volume:

$$V_S = V - V_B \tag{D.9}$$

where V_S is the steam volume, V is the total volume and V_B is the biomass volume. Steam volumes are gathered in vector v:

$$v = \begin{bmatrix} V_{PP}^{S} & V_{TR}^{S} \end{bmatrix}^{\mathrm{T}} \tag{D.10}$$

Steam densities from both containers are placed into the following vector:

$$\rho = \begin{bmatrix} \rho_{PP}^{S} & \rho_{TR}^{S} \end{bmatrix}^{\mathrm{T}} \tag{D.11}$$

Steam density is found by dividing the mass of steam by volume. For example, the steam density in the thermal reactor can be computed in the following way:

$$\rho_2 = \frac{x_3}{v_2} \tag{D.12}$$

where index 2 and 3 selects the corresponding element from vector ρ, x or v.

All pressure variables are gathered in vector p:

$$p = \begin{bmatrix} P_{FS} & P_{PP} & P_{TR} & P_A \end{bmatrix}^{\mathrm{T}} \tag{D.13}$$

where P_A is the outlet pressure of the evacuation valve, which is, in fact, atmospheric pressure.

In case of saturated steam, one steam variable is sufficient to determine any of the other variables. In case of wet or superheated steam, 2 steam variables are necessary to determine its state. Steam properties are inferred from the IAPWS IF97 standard. The fresh steam is saturated and its enthalpy is found from its pressure, which is directly measured. The reactor pressure is determined from enthalpy and density while the temperature is found from enthalpy and pressure. Figure D.3 displays these steam properties. Pressure has a rather linear dependence with respect to density and enthalpy (left plot). Nonlinearities are spotted in temperature variations with respect to pressure in the wet steam region and in the saturation zone (right plot).

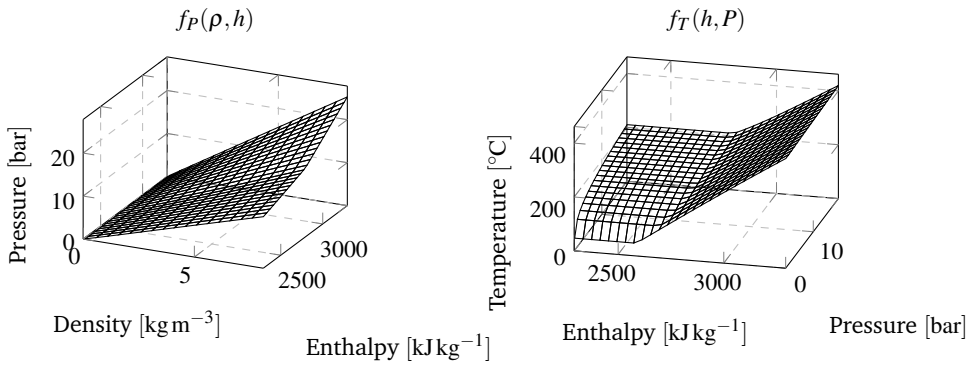


Figure D.3: Steam functions - IAPWS IF97 standard. Left plot shows steam pressure as a function of density and enthalpy while right plot displays steam temperature as a function of enthalpy and pressure.

Considering the notations introduced above, the state derivatives from (D.1a), (D.1c), (D.1b) and (D.1d) become:

$$\dot{x}_1 = f(u_1, p_3, p_2) - f(u_2, p_2, p_4) \tag{D.14a}$$

$$\dot{x}_2 = \frac{1}{x_1}\left[f(u_1, p_3, p_2)\,(x_4 - x_2)\right] \tag{D.14b}$$

$$\dot{x}_3 = f(u_3, p_1, p_3) - f(u_1, p_3, p_2) - d_1 \tag{D.14c}$$

$$\dot{x}_4 = \frac{1}{x_3}\left[f(u_3, p_1, p_3)\,\{g(p_1) - x_4\}\right] \tag{D.14d}$$

$$\dot{x}_5 = \tilde{u}_1 - \tilde{u}_2 \tag{D.14e}$$

$$\dot{x}_6 = \tilde{u}_2 - \tilde{u}_3 \tag{D.14f}$$

where f is the function describing the flow through a steam valve shown in Equation (D.4) and g is the pressure-enthalpy function for saturated steam as found in the IAPWS IF97 standard. States x_{1-4} are coupled with states $x_{5,6}$ through the following equations:

$$p_2 = f_P(\rho_1, x_2) \tag{D.15a}$$

$$p_3 = f_P(\rho_2, x_4) \tag{D.15b}$$

$$\rho_1 = \frac{x_1}{v_1} \tag{D.15c}$$

$$\rho_2 = \frac{x_3}{v_2} \tag{D.15d}$$

$$v_1 = V_{PP} - x_5\rho_B \tag{D.15e}$$

$$v_2 = V_{TR} - x_6\rho_B \tag{D.15f}$$

where ρ_B is the soaked biomass density set to $1000\,\mathrm{kg\,m^{-3}}$.

The model outputs are gathered in vector y and comprise the pressure and temperature in both process units:

$$y = [p_{PP} \quad p_{TR} \quad T_{PP} \quad T_{TR}]^{\mathrm{T}} \tag{D.16}$$

where T_{PP} and T_{TR} are calculated as:

$$y_3 = f_T(x_2, p_2) \tag{D.17a}$$

$$y_4 = f_T(x_4, p_3) \tag{D.17b}$$

Overall, a nonlinear model with 6 states, 3 inputs and 4 outputs is obtained.

D.3.4 Open Loop Simulation

An open loop simulation is prepared in order to observe the system response. The simulation scenario is synthesized in Table D.2.

Table D.2: Open loop simulation scenario.

Time [s]	Q^{FS} [%]	Q^{RS} [%]	Q^{E}_{PP} [%]
0	0	0	0
10	100	0	0
30	0	0	0
50	0	100	0
70	0	0	0
90	0	0	100
110	0	0	0

The reactor fresh steam valve is fully opened at time $t = 10\,\mathrm{s}$ for $20\,\mathrm{s}$. The recycle steam valve is fully opened at time $t = 50\,\mathrm{s}$ for another $20\,\mathrm{s}$. The particle pump evacuation valve is opened at time $t = 90\,\mathrm{s}$ for the last $20\,\mathrm{s}$ of the simulation. The results are plotted in Figure D.4. Pressurization of the thermal reactor takes a much longer time than the particle pump because it has a larger volume. The reactor pressure resembles a second order system response while the temperature could be approximated by a first order response. The particle pump pressurizes in only several seconds and the pressure in both the reactor and the particle pump equalize. There is a small drop in pressure in the reactor because steam is extracted in order to pressurize the particle pump. These disturbances need to be rejected and ensure a steady temperature in the process.

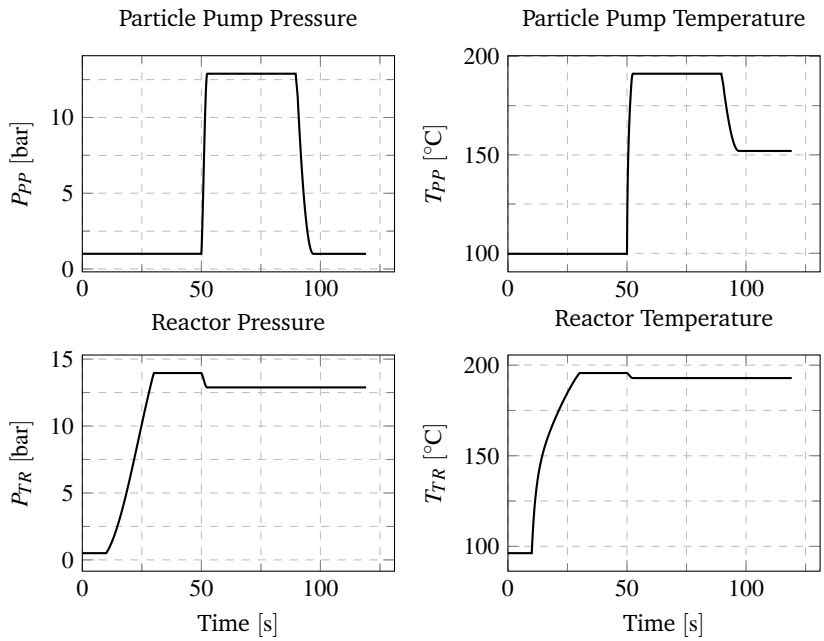


Figure D.4: Open loop simulation. The top plots show the pressure and temperature in the particle pump while the bottom plots display the pressure and temperature in the thermal reactor.

D.4 Control Design

A control strategy is built based on the novel L1 adaptive controller. A nonlinear feed-forward term is also added to the control law based on the measurement or estimation of the recycle steam flow.

Pressurization of the particle pump and its evacuation are achieved by fully opening the recycle and the evacuation valve, respectively. No feedback controllers are required for these actions. The sequence of openings and closings of these valves is determined based on a reference signal R^P_{PP} resembling a square wave with a period τ_{PP}. When the reference signal is 1, the recycle steam valve fully opens and the evacuation valve is closed. When the reference signal is 0, the recycle valve closes and the evacuation valve opens.

D.4.1 Feed-forward Calculation

The flow of recycled steam can be estimated using the valve model formulated in section D.3.2 or directly measured with a flow sensor. It is desired to find the stroke of the thermal reactor fresh steam valve that would compensate the recycled steam.

The stroke of the thermal reactor fresh steam valve is different than the stroke of the recycle valve because the inlet and outlet pressures are different. From Equation (D.4) the valve characteristic K_V can be determined:

$$K_V(t) = \begin{cases} \frac{Q^{RS}(t)}{22.4 \cdot \sqrt{(p_i - p_o)p_o}} & , p_o > \frac{p_i}{2} \\ \frac{Q^{RS}(t)}{11.2 \cdot p_i} & , p_o < \frac{p_i}{2} \end{cases} \tag{D.18}$$

The valve stroke S can be expressed as a function of the valve characteristic K_V by conducting the inverse of function (D.3). However, for simplification, a linear characteristic is used for this computation:

$$S_{FF}(K_V(t)) = \frac{100 K_V(t)}{0.05} \tag{D.19}$$

where 0.05 is the steam volumetric flow (in $\mathrm{m^3\,s^{-1}}$) when the valve is 100 % open with 1 bar pressure drop.

The stroke S_{FF} is fed-forward through a trust gain K_{FF}. The feed-forward control signal becomes:

$$u_{FF}(t) = K_{FF} S_{FF}(t) \tag{D.20}$$

In this application, the trust gain K_{FF} is set to 1 but could be reduced for a more conservative feed-forward action.

D.4.2 L1 Adaptive Control

Figure D.5 shows the system in closed loop with an L1 adaptive output feedback controller. The pressurization of the particle pump is shown in the upper part of the figure. There is no feedback action because the pressure in the particle pump should equalize the pressure in the reactor and this objective can be achieved by fully opening the recycle steam valve. The L1 output feedback control strategy is designed for temperature regulation and only one measurement is considered, i.e. T_{TR}. The manipulated variable is the fresh steam valve stroke, i.e. S_{TR}^{FS}. Therefore, the control object becomes a SISO model.

The L1 output feedback controller consists of an output predictor, an adaptation law and a control filter $C(s)$ [10, 124]. The classical model reference adaptive controller implies a compromise between adaptation and robustness. Moreover, there are no trivial ways of finding a suitable adaptation gain. The L1 adaptive controller separates robustness from fast adaptation by introducing a filter $C(s)$ in the control channel. The analysis of the new controller i.e. the computation of the uniform bounds on outputs and control signals, is performed using the L1 norm, hence the name of L1 adaptive controller.

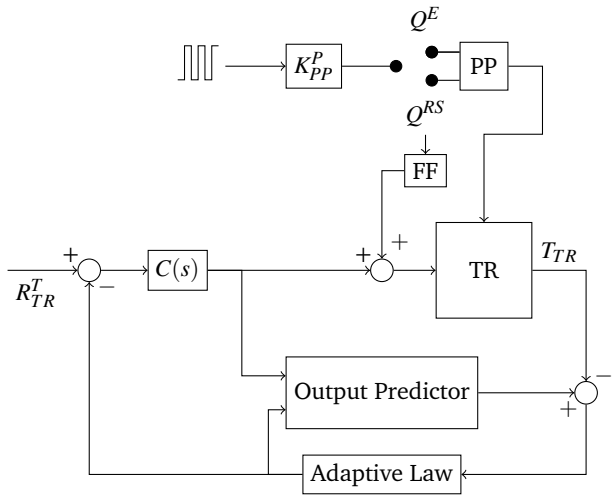


Figure D.5: L1 adaptive output feedback control.

The open loop process can be expressed as follows [10]:

$$y_4(s) = A(s)\{u_3(s) + \tilde{d}(s)\} \tag{D.21}$$

where $y_4(s)$ is the reactor temperature, $A(s)$ is an unknown transfer function, $u_3(s)$ is the fresh steam valve opening and $\tilde{d}(s)$ lumps all the uncertainties and disturbances that affect $A(s)$. Transfer function $A(s)$ can be approximated as the linearized model around a nominal temperature, e.g. 195 °C, resulting the following general structure:

$$A(s) = \frac{K(s+z_1)}{(s+p_1)(s+p_2)} \tag{D.22}$$

where K is the process gain, z_1 is a stable zero, and p_1 and p_2 are real stable poles. Figure D.6 shows the placement of poles p_1 and p_2, and zero z_1 as functions of load and fresh steam valve stroke. The plot was generated for 25 %, 50 % and 75 % load and for 0 %, 25 %, 50 %, 75 % and 100 % valve strokes. The plot tells that, at low load the zero is canceled by a pole and the resulting system resembles a first order response but as the load and the valve stroke increase, not only the system dynamics become faster as its poles get more negative but also the zero detaches from the pole.

The output predictor contains the desired closed loop dynamics and is chosen as a first order system due to the nature of transfer function $A(s)$:

$$\dot{\hat{y}}_4(t) = -m\hat{y}_4(t) + m\{u_3(t) + \hat{\sigma}(t)\} \tag{D.23}$$

where $\hat{y}_4(t)$ is the reactor temperature estimation, m is the pole of the closed loop system, $u_3(t)$ is the fresh steam valve stroke and $\hat{\sigma}(t)$ is the estimation of all uncer-

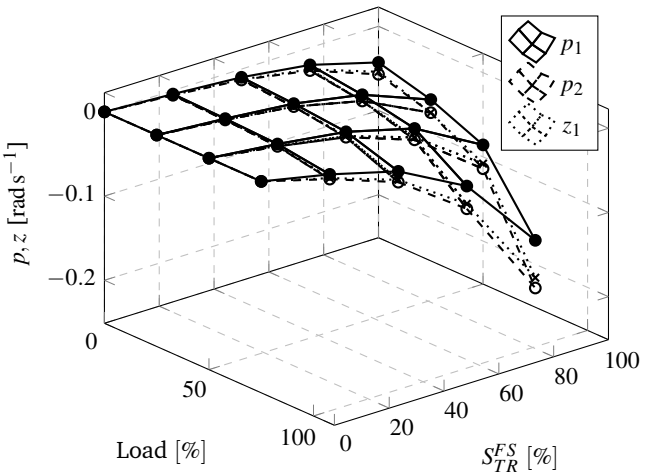


Figure D.6: Poles and zero migration as a function of load and recycle steam valve stroke.

tainties and unmodeled dynamics. In frequency domain, the model reference system is denoted as:

$$M(s) = \frac{m}{s+m} \tag{D.24}$$

System (D.21) can be rewritten in terms of the model reference system [10]:

$$y_4(s) = M(s)\{u_3(s) + \sigma(s)\} \tag{D.25}$$

where $\sigma(s)$ is identified as:

$$\sigma(s) = \frac{\{A(s) - M(s)\}u_3(s) + A(s)\tilde{d}(s)}{M(s)} \tag{D.26}$$

The idea is to cancel uncertainties $\sigma(s)$ with the help of the control signal $u_3(s)$. Therefore, the control signal is defined as [10]:

$$u_3(s) = C(s)\{r(s) - \hat{\sigma}(s)\} \tag{D.27}$$

where $C(s)$ is chosen in this application as a first order filter:

$$C(s) = \frac{c}{s+c} \tag{D.28}$$

If Equation (D.27) is substituted in (D.25) then:

$$y_4(s) = M(s)C(s)r(s) + M(s)\{\sigma(s) - C(s)\hat{\sigma}(s)\} \tag{D.29}$$

If $\sigma(s)$ is perfectly estimated then $\hat{\sigma}(s) = \sigma(s)$ and the disturbances will be rejected only in the bandwidth of $C(s)$.

In [10] the adaptation law is implemented as a projection algorithm. In this application, because there is only one parameter to estimate, the adaptation law is simplified to an integrator:

$$\dot{\hat{\sigma}}(t) = -\Gamma \tilde{y}_4(t) \tag{D.30}$$

where $\tilde{y}_4(t) = \hat{y}_4(t) - y_4(t)$ is the estimation error of the output and Γ is the integrator gain.

The resulting L1 adaptive output feedback controller has 3 tuning parameters, i.e. the desired closed loop dynamics m, the adaptation gain Γ and the eigenfrequency of the control filter c.

The reference model and the control signal filter can be designed systematically [125]. Assuming perfect knowledge of disturbances, an ideal system $y_4(s)$ can be built and used for tuning [125]:

$$y_4(s) = H(s)C(s)r(s) + H(s)\{1 - C(s)\}\tilde{d}(s) \tag{D.31}$$

where $H(s)C(s)$ is the transfer function from $r(s)$ to $y_4(s)$ and $H(s)$ is defined as:

$$H(s) = \frac{A(s)M(s)}{C(s)A(s) + \{1 - C(s)\}M(s)} \tag{D.32}$$

Parameters m and c must be chosen such that $H(s)$ is stable and the following L1 norm holds [10]:

$$||G(s)||_{\mathscr{L}_1} L < 1 \tag{D.33}$$

where $G(s) = H(s)\{1 - C(s)\}$ is the transfer function from $\tilde{d}(s)$ to $y_4(s)$ and L is the Lipschitz constant required to guarantee BIBO stability (Lemma 4.1.1 in [10]).

In this application, parameters m, c and Γ are found by minimizing the integral absolute error (IAE) function:

$$\min_{m,c,\Gamma} \int_0^\infty |r(t) - y_4(t)| \tag{D.34}$$

where $r(t) - y_4(t)$ is the tracking error and $y_4(t)$ is the output of the nonlinear model.

A numerical search procedure is instantiated to obtain parameters m, c and Γ. The simulation scenario for minimizing the objective function contains the normal operation cycle, i.e. the particle pump extracting steam from the reactor, which causes a disturbance in the reactor temperature. Therefore, the controller is optimized for disturbance rejection. The objective function (D.34) cannot be minimized analytically due to the complexity of the nonlinear model. The numerical search procedure performs various simulations in closed loop, for each simulation evaluating the IAE function from (D.34). The search algorithm is initialized by following the tuning procedure shown in [12]. A temperature reference of 195 °C

and 0 % biomass load are held in all simulations. The numerical search procedure found the following parameters:

$$m = 0.07 \qquad c = 20 \qquad \Gamma = 3135 \tag{D.35}$$

With these optimal parameters, transfer function $H(s)$ becomes:

$$H(s) = 0.018\frac{s+20}{(s+5.3)(s+0.07)} \tag{D.36}$$

which has stable poles, as required by the design specifications. Transfer function $G(s)$ is:

$$G(s) = 0.018\frac{s}{(s+0.07)(s+5.3)} \tag{D.37}$$

The L1 norm of $G(s)$ is computed and the Lipschitz constant is found:

$$L = 310.64 \tag{D.38}$$

which ensures BIBO stability of the closed loop system according to Lemma 4.1.1 from [10].

The tracking error between a real and an ideal system, which assumes perfect knowledge of the disturbances, is uniformly bounded with respect to a constant proportional to $1/\sqrt{\Gamma}$ [10]. The larger Γ is, the better is the performance. Therefore, a high adaptation gain Γ is desired. At the same time, the stability and dynamics of $\hat{\sigma}$ depend on Γ. The transfer function from $r(s)$ and $d(s)$ to $\hat{\sigma}(s)$ is [125]:

$$\hat{\sigma} = F(s)[C(s)\{A(s) - M(s)\}r(s) + A(s)d(s)] \tag{D.39}$$

where $F(s)$ is identified as:

$$F(s) = \frac{1}{\frac{s}{\Gamma} + C(s)A(s) + \{1 - C(s)\}M(s)} \tag{D.40}$$

Γ should be chosen such that $F(s)$ is stable. Considering the optimal parameters from (D.35), the following poles of $F(s)$ are found:

$$\begin{bmatrix} -5.04 \pm 9.26i & -10 & -0.07 & -0.011 \end{bmatrix} \tag{D.41}$$

which are all stable. Therefore, the adaptation algorithm is stable.

D.5 Benchmark Tests

The L1 adaptive controller is tested in three scenarios that are described below. The feed-forward term is always enabled and the reactor biomass load is set to 25 %.

The first scenario corresponds to the normal operation cycle, which assumes pressurization of the particle pump with recycled steam every 180 s. The controller has to keep the reactor temperature at an optimal level, e.g. 195 °C. Condensation of steam cannot be directly measured and is treated as unmeasured disturbance. Steam condensation is modeled as normally distributed white noise with mean m_C and standard deviation σ_C:

$$Q_{TR}^{C} \in N(m_C, \sigma_C) \tag{D.42}$$

where $m_C = 2\,\mathrm{kg\,s^{-1}}$ and $\sigma_C = 0.3\,\mathrm{kg\,s^{-1}}$.

The second scenario treats measurement noise, which affects the computation of the feed-forward term. The measurement noise specifications are:

$$Q_{RS}^{M} \in N(m_M, \sigma_M) \tag{D.43}$$

where $m_M = 0\,\mathrm{kg\,s^{-1}}$ and $\sigma_M = 0.5\,\mathrm{kg\,s^{-1}}$.

The temperature reference signal changes during production mode typically when the refinery switches to another type of biomass. Therefore, the temperature controller is also tested with respect to reference step changes in the last simulation scenario.

D.6 Results

The scenario comprising a normal operation with condensation disturbances can be observed in Figure D.7. The pressurization of the particle pump is shown in the top left subplot. Two cycles were captured in the plot. The thermal reactor pressure and temperature are displayed in the bottom subplots. Naturally, when the particle pump starts to inhale steam from the thermal reactor, the temperature inside the reactor drops as a consequence of the pressure drop. However, the controller with the feed-forward term is able to reject the disturbance effectively leading to an unnoticeable change in the reactor temperature and pressure as illustrated in the bottom subplots of the figure. The control effort is shown in the top right subplot and is translated into mass flow of fresh steam. It has a non-zero steady-state value due to a non-zero mean value of the condensation disturbance.

In the second simulation scenario, the feed-forward term is considered partially non-reliable and a large noise is injected in the measurement of the recycle steam flow. The results can be observed in Figure D.8. The L1 controller tolerates feed-forward noise effectively. The reactor pressure is not as white as in the first scenario but it translates to negligible variations in the reactor temperature.

The reference tracking case is treated in Figure D.9. The feed-forward term has no effect since the particle pump is stopped when changing the reactor temperature.

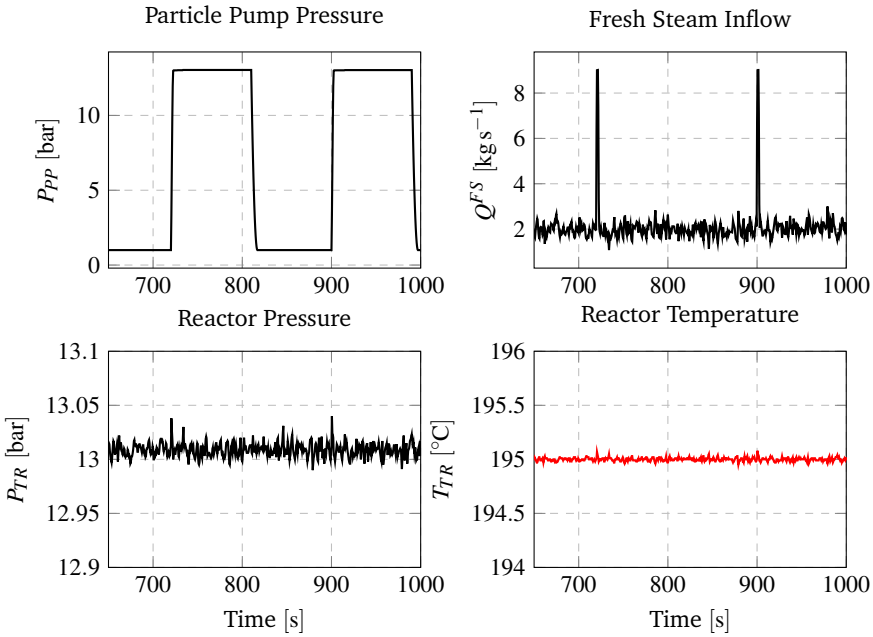


Figure D.7: Normal operation cycle with condensation noise.

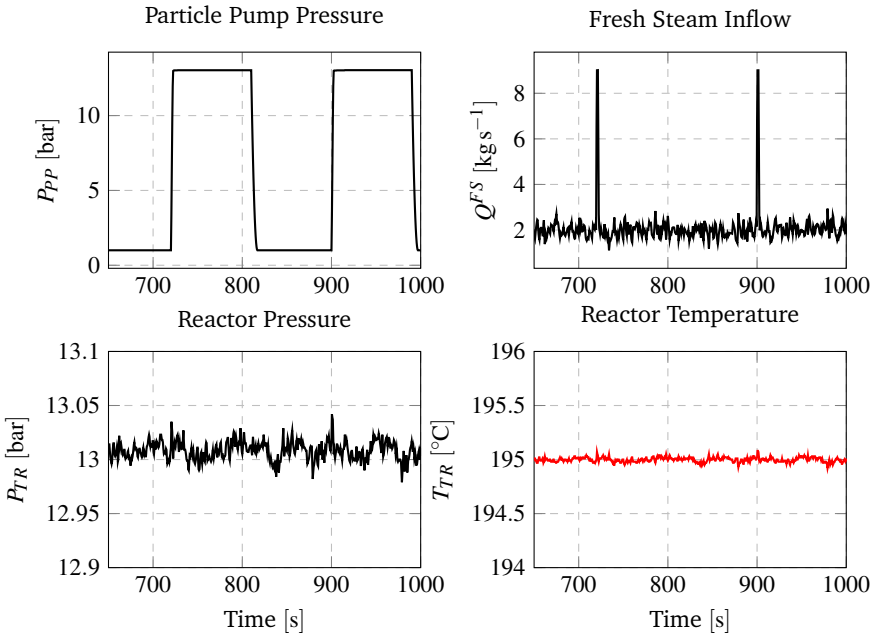


Figure D.8: Normal operation cycle with measurement noise.

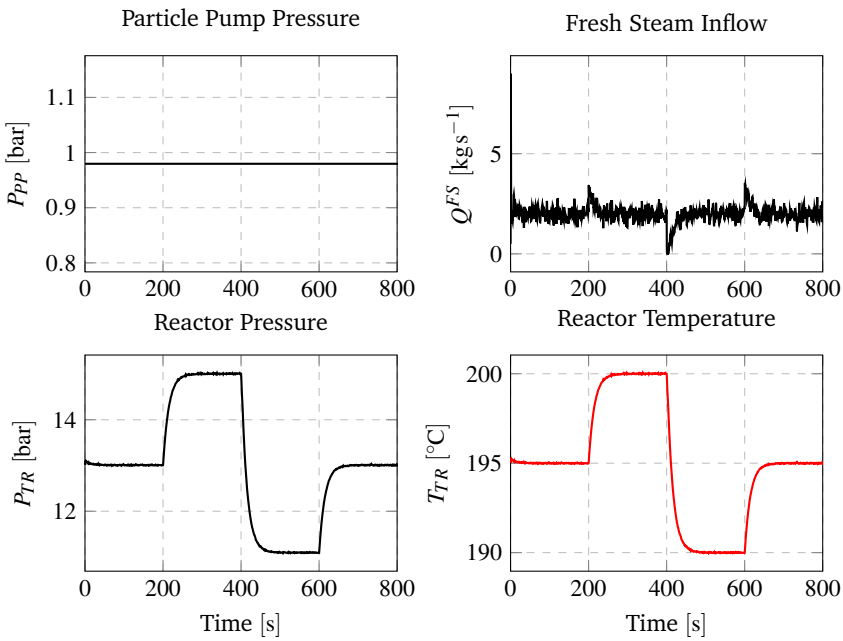


Figure D.9: Reference tracking with condensation noise.

Cooling the reactor is performed by interrupting the supply of fresh steam. The L1 adaptive controller has good performances changing the temperature in less than 1 min, which is satisfactory since temperature set point changes do not occur frequently in production mode.

D.7 Conclusions

For the first time, to our knowledge, a steam layer model has been formulated for a particle pump connected to a thermal reactor in biomass pretreatment. The process model embeds nonlinear valve characteristics and steam states dependencies, e.g. pressure, temperature, enthalpy and density.

The developed control strategy based on the novel L1 adaptive output feedback controller proved to have very good performances in all simulation scenarios. Measurement noise that affects the feed-forward term and condensation disturbances were handled satisfactory. The application of such a controller to regulate the temperature in a biomass pretreatment reactor has not been reported earlier in the literature.

The last achievement of the paper is the tuning method of the L1 adaptive controller. It proved successfully to formulate the tuning of the controller as an

optimization problem. The integral absolute error (IAE) has been used as an objective function and the determined controller parameters proved to give satisfactory results in simulation.

The temperature controller developed in this paper is part of an inner control loop layer of a biorefinery. The temperature setpoint will be given by an outer control loop layer that is using the pretreated biomass composition as feedback. The objective is to obtain a steady pretreatment process and any variations in biomass composition or type can be rejected by changing the temperature in the reactor.

Paper E

Modeling and L1 Adaptive Control of pH in Bioethanol Enzymatic Process

Remus Mihail Prunescu[1], Mogens Blanke[*1], Gürkan Sin[2]

[1]Department of Electrical Engineering, Automation and Control Group, Technical University of Denmark, Elektrovej Building 326, 2800, Kgs. Lyngby, Denmark

[2]CAPEC-PROCESS, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads Buildings 227 and 229, 2800, Kgs. Lyngby, Denmark

Abstract:
The enzymatic process is a key step in second generation bioethanol production. Pretreated biomass fibers are liquefied with the help of enzymes to facilitate fermentation. Enzymes are very sensitive to pH and temperature and the main control challenge in the nonlinear process is to ensure minimum deviations from the optimal pH level. This article develops a mathematical model for the pH, which has not been reported earlier for this particular process. The new model embeds flow dynamics and pH calculations and serves both for simulation and control design. Two control strategies are then formulated for pH level regulation: one is a classical PI controller; the other an L1 adaptive output feedback controller. Model-based feed-forward terms are added to the controllers to enhance their performances. A new tuning method of the L1 adaptive controller is also proposed. Further, a new performance function is formulated and tailored to this type of processes and is used to monitor the performances of the process in closed loop. The L1 design is found to outperform the PI controller in all tests.

[*]Principal corresponding author. Tel.: +45 45253565; E-mail: mb@elektro.dtu.dk

E.1 Introduction

Bioethanol is thought to become the primary renewable liquid fuel [1, 126] and extensive endeavors have been conducted to make the production process feasible on a large scale [3]. In order to reduce operating costs, bio-refineries are integrated with power plants following the IBUS principle [4, 55]. The conversion of lignocellulosic biomass to ethanol is performed only with steam and enzymes, therefore categorizing the technology as a green process.

In a second generation bioethanol production process, the enzymatic liquefaction step prepares the pretreated biomass for fermentation [4, 55]. In the pretreatment phase, the biomass is soaked with acetic acid and then pretreated with steam. Studies show that a steam/acetic acid combination improves the pretreatment process [121] and it will be considered that the stream of fibers is mildly acidic, thus lowering the pH level. Acetic acid is also produced in reduced quantities during the pretreatment stage.

The enzyme activity is influenced both by pH and temperature. Enzyme activity versus pH follows a certain bell-shaped curve [123]. Optimal pH activity differs for each enzyme type. For example, in the case of Accellerase TRIO, a pH level of 4.8 should be set [2]. Another type of enzymes, i.e. Cellic CTec3 produced by Novozymes, requires a pH of 5.0 [3]. A small deviation from the pH optimal value, e.g. 0.2, can cause a significant drop in the process efficiency, e.g. $20\,\%$. Therefore, the main control challenge in such a process is to minimize deviations from the optimal pH level.

Controlling the pH has been the topic of many research activities and there are many generic solutions in the literature. A comprehensive review of the existing generic control strategies can be found in [127]. Control strategies vary from pure feed-forward control, where the inflow of hydroxide is manually adjusted by an operator, to more sophisticated schemes like adaptive fuzzy control [128], nonlinear adaptive control [129, 130] or model predictive control [131, 132]. The most common application of pH control is the neutralization process but, in an enzymatic liquefaction process, it is desired to keep the pH level somewhere in the range 4-6 depending on the enzyme type.

To our knowledge, pH modelling for an enzymatic liquefaction process has not been conducted earlier. Therefore, a process model is first formulated. Two control schemes are then designed, a PI controller and an L1 adaptive output feedback controller. Classical PI control is wide spread in industry and is used as a reference

[2] Accelerase® TRIO™ datasheet from Genencor.com

[3] Cellic CTec3 datasheet from Novozymes.com

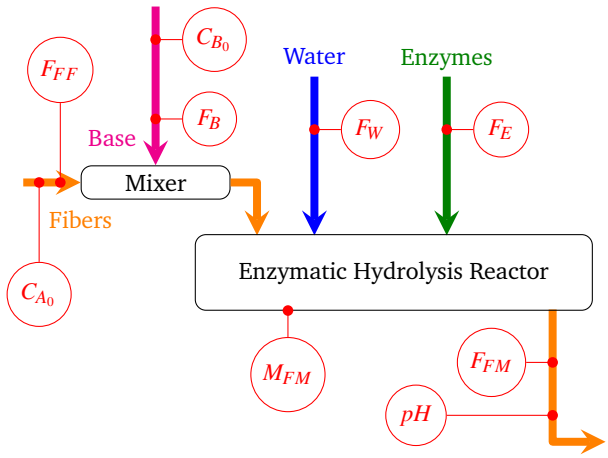


Figure E.1: Process diagram with instrumentation.

for control performance. The L1 adaptive controller represents a novelty in control theory and has never been applied to pH control of liquefaction processes. Major disturbances will be identified and model based feed-forward terms will be added to the controllers in order to improve their action. Also, a new tuning method of the L1 controller is proposed as an enhancement of the method developed in [125]. Finally, the controllers are tested on large scale benchmark scenarios that cover both reference tracking and disturbance rejection cases. A performance function is formulated based on the enzymatic activity curve and the controllers are compared using the enzymatic efficiency associated with their action.

E.2 Process Description

A generic enzymatic liquefaction process is illustrated in Figure E.1.

The pretreated biomass was decomposed into fibers in the pretreatment stage and arrives in the liquefaction reactor having a high dry matter content TS typically between $25-30\,\%$. A high dry matter is required in order to make the technology cost-effective but it cannot increase indefinitely due to mixing technical problems that may appear [55, 133]. A sample of fibers is extracted automatically and an NIR analysis is performed in order to determine its chemical composition, including the concentration of acetic acid denoted as $[\mathrm{AcH}]_0$ or C_{A_0}, which is of interest in this case. The mass flow of fibers is also directly measured and denoted as F_{FF}.

pH control is achieved by pumping a strong base i.e. sodium hydroxide, NaOH, into the fiber fraction before entering the liquefaction tank. The inflow of base is measured as F_B and its concentration is $10\,\mathrm{wt}\%$.

Enzymes are pumped into the reactor proportional to cellulose quantity and the inflow of enzymes is measured as F_E. A quantity of water F_W is also poured into the tank. The weight of the tank, the outflow and the pH value are directly measured as M_{FM}, F_{FM}, and pH, respectively. Nominal values of the inflows, acetic acid concentration, tank load and total solids are presented in Table E.1. The data refer to a bioethanol plant that was designed to handle $1000\,\mathrm{kg\,h^{-1}}$ of fibers. The enzymatic liquefaction tank has a nominal load capacity of 5000 kg of fiber mash and the concentration of acetic acid varies around $5\,\mathrm{g\,kg^{-1}}$ depending on the pretreatment process parameters.

Table E.1: Nominal operation of a hydrolysis process.

F_{FF}	1000	$\mathrm{kg\,h^{-1}}$	$[\mathrm{AcH}]_0$	5	$\mathrm{g\,kg^{-1}}$
F_E	20	$\mathrm{kg\,h^{-1}}$	M_{FM}	5000	kg
F_W	80	$\mathrm{kg\,h^{-1}}$	TS	25	%

E.3 Control Challenge

Several theoretical enzymatic activity bell-shaped curves are shown in Figure E.2. The optimal pH level corresponds to an enzymatic activity of 100 %. In case of disturbances that occur in the pretreatment process, the acetic acid concentration in the inflow changes and affects the pH level. A small deviation from the optimal pH level can cause a significant drop in the enzymatic activity. This means that the quality of the outflow drops. An increase in enzyme quantity is necessary to meet the same quality constraints on the outflow but enzymes are very expensive. It is a lot cheaper to properly control the pH level with a base.

Another control challenge arises from the large scale nature of the process. The pH sensor is positioned on the outflow of the tank because it is easier to measure the pH level on a liquefied substance. The flow dynamics of the reactor are slow and a large quantity of fibers can be compromised due to a small deviation in the pH level. Therefore, a model and feed-forward terms for the controllers are necessary to obtain a high performance control strategy.

E.4 Process Model

The pH calculation is derived using the classical physico-chemical approach, which considers a set of weak acid/base equilibrium in liquid phase [112] and gas-liquid CO_2 stripping process [113]. It is assumed that CO_2 is not produced in the enzymatic liquefaction process but traces of bicarbonate can exist in the inflows due

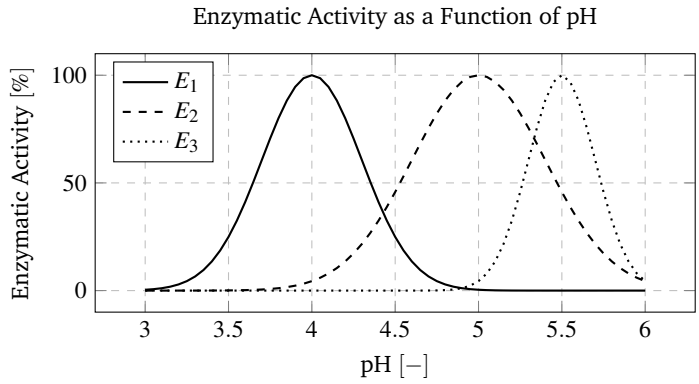


Figure E.2: Different enzymatic activity bell-shaped curves.

to upstream subprocesses or process water utilization. The process model captures flow dynamics and considers the mixture in chemical equilibrium at any given time.

E.4.1 pH Calculation

In total, the model has 6 weak acid/base equilibrium equations and CO_2 stripping. It is also assumed that production and consumption of ions are negligible from enzymatic reactions.

Acetic acid is a weak acid and partially dissociates into Ac^- and H^+ with an equilibrium constant K_A (E.1a). Sodium hydroxide is a strong base and fully dissociates into Na^+ and OH^- (E.1b). Ac^- and Na^+ combine to form the salt sodium acetate NaAc (E.1c). Water self-ionizes with an equilibrium constant K_W (E.1d). The liquid phase is not pure and is considered to contain CO_2, which is the cause of a buffer formation that affects the pH level. Carbon dioxide forms carbonic acid H_2CO_3 in water, which dissociates with an equilibrium constant K_{C_1} into bicarbonate HCO_3^- and H^+ as in (E.1e). The bicarbonate continues to decompose forming carbonate CO_3^{2-} and H^+ with an equilibrium constant K_{C_2} like in (E.1f).

$$\mathrm{AcH} \rightleftharpoons \mathrm{Ac^- + H^+} \tag{E.1a}$$

$$\mathrm{NaOH} \longrightarrow \mathrm{Na^+ + OH^-} \tag{E.1b}$$

$$\mathrm{Ac^- + Na^+} \longrightarrow \mathrm{NaAc} \tag{E.1c}$$

$$\mathrm{H^+ + OH^-} \rightleftharpoons \mathrm{H_2O} \tag{E.1d}$$

$$\mathrm{CO_2 + H_2O} \longrightarrow \mathrm{H_2CO_3} \rightleftharpoons \mathrm{HCO_3^- + H^+} \tag{E.1e}$$

$$\mathrm{HCO_3^-} \rightleftharpoons \mathrm{CO_3^{2-} + H^+} \tag{E.1f}$$

Three variables are defined:

$$[\mathrm{A_T}] = [\mathrm{AcH}] + [\mathrm{Ac^-}] \tag{E.2a}$$
$$[\mathrm{B_T}] = [\mathrm{Na^+}] + [\mathrm{NaOH}] \tag{E.2b}$$
$$[\mathrm{C_T}] = [\mathrm{CO_2}] + [\mathrm{HCO_3^-}] + [\mathrm{CO_3^{2-}}] \tag{E.2c}$$

The presence of other acids like lactic acid or succinic acid was also recorded but in negligible amounts and there might be other buffers in the stream. Therefore, the concentration of all unmodeled ions will be lumped into a single variable $[\mathrm{Z^-}]$ and the charge balance is then formulated:

$$\begin{aligned}&[\mathrm{H^+}] - [\mathrm{Z^-}] + [\mathrm{Na^+}] - \\ &-[\mathrm{OH^-}] - [\mathrm{Ac^-}] - [\mathrm{HCO_3^-}] - 2[\mathrm{CO_3^{2-}}] = 0\end{aligned} \tag{E.3}$$

Each term from the charge balance can be found as a function of variables $[\mathrm{A_T}]$, $[\mathrm{B_T}]$, $[\mathrm{C_T}]$ or $[\mathrm{H^+}]$ and equilibrium constants K_A, K_W, K_{C_1} and K_{C_2}. Therefore, $[\mathrm{H^+}]$ can be determined by finding the real positive zero of (E.3). Afterwards, the pH level is computed using its definition:

$$\mathrm{pH} \stackrel{def}{=} -\log_{10}[\mathrm{H^+}] \tag{E.4}$$

There might be an offset between the estimated and measured pH level. The difference can be canceled by adjusting the concentration of $\mathrm{Z^-}$ i.e. C_{Z_0}. An online estimation of $[\mathrm{Z^-}]$ could be implemented in reality.

E.4.2 State Space Model

In this section, the flow dynamics and the pH calculation are embedded into a single model that has the structure from Figure E.3. It is important to conduct all pH calculations only on the liquid part of the mixture. All notations are gathered in Table E.2 with their measuring units.

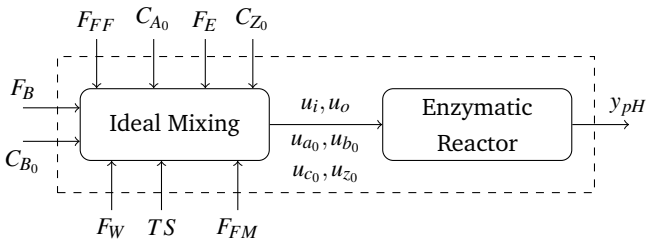


Figure E.3: Process flow scheme.

Perfect and instantaneous mixing of the tank is assumed and this is represented as an *Ideal Mixing* block, which combines all inflows into a single input u_i without solids. The outflow without solids is denoted as u_o. The inflow concentrations u_{a_0}, u_{b_0} and u_{c_0} include a measuring unit transformation and are expressed in $\mathrm{mol\,L^{-1}}$.

The state variables of the dynamic model are defined as:

$$x_z = [\mathrm{Z^-}] \tag{E.5a}$$

$$x_{a_t} = [\mathrm{A_T}] \tag{E.5b}$$

$$x_{b_t} = [\mathrm{B_T}] \tag{E.5c}$$

$$x_{c_t} = [\mathrm{C_T}] \tag{E.5d}$$

and x_m, which is the mass of liquid inside the reactor.

Two more algebraic variables are also defined:

$$x_h = [\mathrm{H^+}] \tag{E.6a}$$

$$x_{co_2} = [\mathrm{CO_2}] \tag{E.6b}$$

The system dynamics read as:

$$\frac{dx_m}{dt} = u_i - u_o \tag{E.7a}$$

$$\frac{d(x_m x_z)}{dt} = u_i u_{z_0} - u_o x_z \tag{E.7b}$$

$$\frac{d(x_m x_{a_t})}{dt} = u_i u_{a_0} - u_o x_{a_t} \tag{E.7c}$$

$$\frac{d(x_m x_{b_t})}{dt} = u_i u_{b_0} - u_o x_{b_t} \tag{E.7d}$$

$$\frac{d(x_m x_{c_t})}{dt} = u_i u_{c_0} - u_o x_{c_t} + r_{ctr} x_m \tag{E.7e}$$

where r_{ctr} is the $\mathrm{CO_2}$ stripping rate [113]:

$$r_{ctr} = k_{ctr}\left(u^*_{co_2} - x_{co_2}\right) \tag{E.8}$$

The dissolved concentration of $\mathrm{CO_2}$ is denoted as $u^*_{co_2}$ and is governed by the Henry law [113].

The charge balance can be rewritten as a polynomial:

$$x_h^5 + p_1 x_h^4 + p_2 x_h^3 + p_3 x_h^2 + p_4 x_h + p_5 = 0 \tag{E.9}$$

where coefficients p_i are found by identification after expanding equation (E.3). The nonlinearity of the process arises from the fact that coefficients p_i are functions of model states and change in time. x_h is then determined as the positive real zero of polynomial (E.9) and the output of the model is defined as:

$$y_{pH} = -\log_{10} x_h \tag{E.10}$$

Table E.2: Explanation of symbols that were used in the process model.

u_i	Liquid part of fibre fraction inflow	$\mathrm{kg\,h^{-1}}$
u_o	Liquid part of fibre mash outflow	$\mathrm{kg\,h^{-1}}$
u_{a_0}	Inflow concentration of acetic acid	$\mathrm{mol\,L^{-1}}$
u_{b_0}	Inflow concentration of sodium hydroxide	$\mathrm{mol\,L^{-1}}$
u_{c_0}	Inflow concentration of carbon dioxide	$\mathrm{mol\,L^{-1}}$
u_{z_0}	Inflow concentration of unmodelled ions	$\mathrm{mol\,L^{-1}}$
x_m	Mass of liquid fibre mash	kg
x_{a_t}	Total molar concentration of acid species	$\mathrm{mol\,L^{-1}}$
x_{b_t}	Total molar concentration of base species	$\mathrm{mol\,L^{-1}}$
x_{c_t}	Total molar concentration of carbonic species	$\mathrm{mol\,L^{-1}}$
x_z	Molar concentration of unmodelled ions	$\mathrm{mol\,L^{-1}}$
x_h	Total molar concentration of H^+	$\mathrm{mol\,L^{-1}}$
x_{co_2}	Total molar concentration of CO_2	$\mathrm{mol\,L^{-1}}$
$u_{co_2^*}$	Molar concentration of CO_2 in the atmosphere	$\mathrm{mol\,L^{-1}}$
k_{ctr}	CO_2 stripping process parameter	$\mathrm{h^{-1}}$
K_W	H_2O dissociation constant	-
K_A	AcH dissociation constant	-
K_{C_1}	CO_2 dissociation constant	-
K_{C_2}	HCO_3 dissociation constant	-
ρ_{FM}	Fiber mash density	$\mathrm{kg\,L^{-1}}$
M_A	Molar mass of AcH	$\mathrm{g\,mol^{-1}}$
M_B	Molar mass of NaOH	$\mathrm{g\,mol^{-1}}$
C_{A_0}	Inflow concentration of AcH	$\mathrm{g\,kg^{-1}}$
C_{B_0}	Inflow concentration of NaOH	$\mathrm{g\,kg^{-1}}$
C_{C_0}	Inflow concentration of CO_2	$\mathrm{g\,kg^{-1}}$
C_{Z_0}	Inflow concentration of Z^-	$\mathrm{mol\,L^{-1}}$
TS	Total solids in inflow stream	%

The numerical values of all model parameters are listed in Table E.3.

Table E.3: Model parameters.

K_W	1×10^{-14}	M_A	$60.05221\,\mathrm{g\,mol^{-1}}$
K_A	1.7378×10^{-5}	M_B	$39.99715\,\mathrm{g\,mol^{-1}}$
K_{C_1}	4.3003×10^{-7}	M_C	$44.01\,\mathrm{g\,mol^{-1}}$
K_{C_2}	4.7995×10^{-11}	C_{A_0}	$5\,\mathrm{g\,kg^{-1}}$
$u^*_{co_2}$	$1.71 \times 10^{-5}\,\mathrm{mol\,L^{-1}}$	C_{B_0}	$100\,\mathrm{g\,kg^{-1}}$
ρ_{FM}	$1.05\,\mathrm{kg\,L^{-1}}$	C_{C_0}	$7.1673 \times 10^{-4}\,\mathrm{g\,kg^{-1}}$
		C_{Z_0}	$0.08\,\mathrm{mol\,L^{-1}}$

E.4.3 Titration Simulation

The model is tested by performing a classical titration simulation, i.e. feeding into the process an inflow of hydroxide resembling a stairway shape with a stair amplitude of $1\,\mathrm{kg\,h^{-1}}$ starting at time 20 h. The concentration of Z^- is disregarded. The results can be observed in Figure E.4.

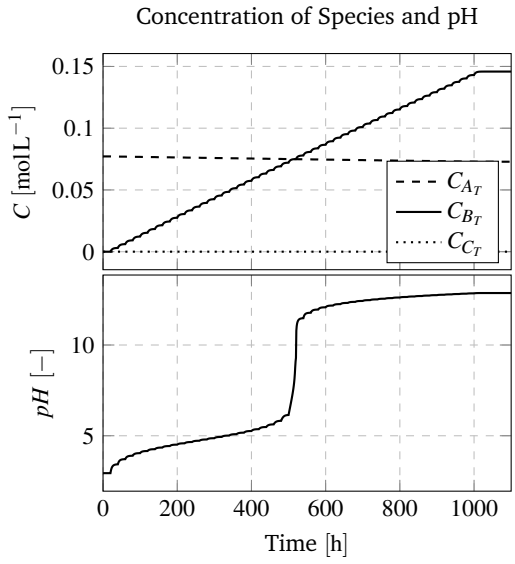


Figure E.4: Titration simulation for verifying the implementation of the model.

The acetic acid and the carbonic acid keep a low pH but after adding a substantial amount of base, approximately at time $t = 500\,\mathrm{h}$, the buffers are depleted and the pH increases to a value greater than 7, which is an expected result.

E.5 Control Design

The enzymatic liquefaction tank is assumed to have a mass controller, which can be easily constructed as a feed-forward strategy combined with feedback.

A bioethanol plant may switch between enzyme types but such an action does not happen often during production mode. Therefore, reference tracking is necessary but emphasis will be placed on disturbance rejection. There are many sources of disturbances that affect the pH level, e.g. pretreatment conditions, which influence the concentration of acetic acid, the presence of other weak acids and negative ions or the imperfect mixing effects, which might be significant since the tank has a large volume and a homogenous environment cannot be guaranteed.

E.5.1 Feed-forward Combined With PI Control

In this section, a classical PI controller combined with a nonlinear feed-forward signal is derived following a traditional design algorithm, i.e. linearization of the process model around the nominal operational point from Table E.1 with $pH = 5$, and derivation of the controller using the Skogestad internal model control (SIMC) approach. The pH level is mainly disturbed by the initial concentration of AcH, which is measured through NIR analysis. A feed-forward term can be created using the nonlinear model to compensate for this type of disturbance. The block diagram of the closed loop system is shown in Figure E.5.

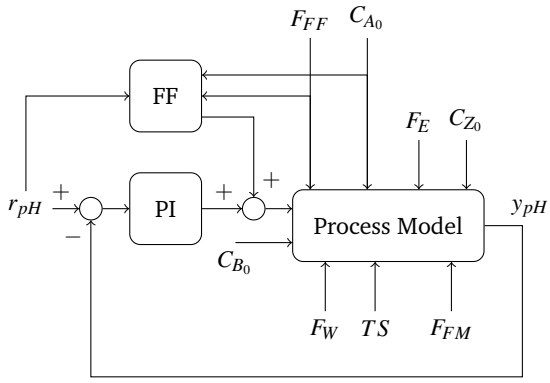


Figure E.5: Closed loop system with a feed-forward and a PI controller.

The PI control law is defined as:

$$u_{PI}(t) = K_P e(t) + K_I \int_0^t e(t) \tag{E.11}$$

where $u_{PI}(t)$ is the feedback contribution to the flow of hydroxide, $e(t)$ is the pH error signal defined as $e(t) = r_{pH}(t) - y_{pH}(t)$, $r_{pH}(t)$ is the pH reference signal, K_P is the proportional gain and K_I is the integral gain. Following SIMC rules [90], K_P and K_I were set to:

$$K_P = 22.8\,\mathrm{kg/h/pHunit} \tag{E.12a}$$

$$K_I = 16.8\,\mathrm{kg/h/pHunit} \tag{E.12b}$$

The feed-forward term is found from the charge balance (E.3) from where $[\mathrm{Na}^+]$ or u_{b_0} is isolated:

$$[\mathrm{Na}^+] = -[\mathrm{H}^+] + [\mathrm{OH}^-] + [\mathrm{Ac}^-] + [\mathrm{HCO_3}^-] + 2[\mathrm{CO_3}^{2-}] + [\mathrm{Z}^-] \tag{E.13}$$

The required concentration of $[\mathrm{H}^+]$ is found from the reference level:

$$[\mathrm{H}^+] = 10^{-r_{pH}} \tag{E.14}$$

and all the other concentrations are derived with the help of $[\mathrm{H}^+]$ and using the steady-state values for $[\mathrm{C_T}]$ and $[\mathrm{A_T}]$. The contribution of the feed-forward term to the total inflow of hydroxide can then be computed. The total flow of base becomes:

$$F_{OH_i} = u_{FF} + u_{PI} \tag{E.15}$$

where u_{FF} is the feed-forward contribution and u_{PI} is computed by the controller using feedback action.

E.5.2 L1 Adaptive Control

The L1 adaptive controller is a modified version of the model reference adaptive controller with a state predictor. In the MRAC architecture, the key to a high performance control strategy is the adaptation gain, which is subject to a compromise and there are no systematic ways of finding an optimal value. A high adaptation gain introduces high-frequency noise in the control channel and the stability margins are affected. In order to separate robustness from adaptation performance, a filter $C(s)$ is introduced in the control channel [10]. The analysis of the new controller i.e. the computation of the uniform bounds on outputs and control signals, is performed using the L1 norm, hence the name of L1 adaptive controller.

The process in open loop can be expressed as follows [10]:

$$y(s) = A(s)\{u(s) + d(s)\} \tag{E.16}$$

where $y(s)$ is the pH level, $A(s)$ is an unknown transfer function, $u(s)$ is the system input and $d(s)$ lumps all the uncertainties and disturbances that affect $A(s)$. Transfer

function $A(s)$ can be approximated as the linearized model around the nominal point from Table E.1 with a base flow set such that $pH = 5$.

The L1 adaptive output feedback controller consists of an output predictor, an adaptation law and a control law [10, 124]. The structure of the closed loop system with this type of controller is presented in Figure E.6.

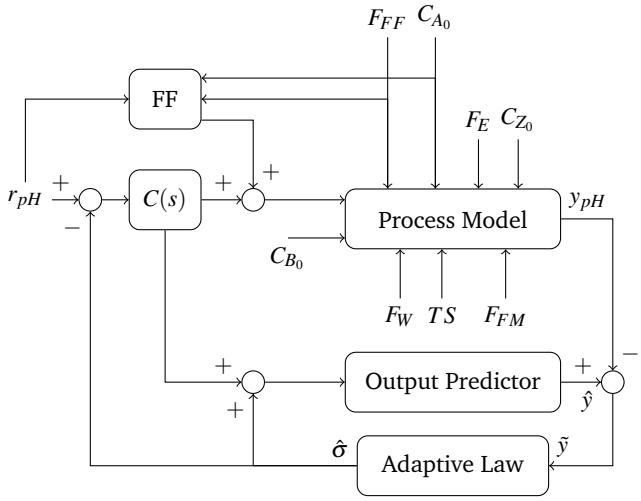


Figure E.6: Closed loop system with an L1 adaptive output feedback controller.

The output predictor is built with the help of a first order model reference system:

$$\dot{\hat{y}}(t) = -m\hat{y}(t) + m\{u(t) + \hat{\sigma}(t)\} \tag{E.17}$$

where $\hat{y}(t)$ is the output estimation, $1/m$ is the desired time constant of the closed loop system, $u(t)$ is the inflow of hydroxide and $\hat{\sigma}(t)$ is the estimation of all uncertainties and unmodelled dynamics. The model reference system is also denoted as:

$$M(s) = \frac{m}{s+m} \tag{E.18}$$

System (E.16) can be rewritten in terms of the model reference system [10]:

$$y(s) = M(s)\{u(s) + \sigma(s)\} \tag{E.19}$$

where $\sigma(s)$ is identified as [10]:

$$\sigma(s) = \frac{\{A(s) - M(s)\}u(s) + A(s)d(s)}{M(s)} \tag{E.20}$$

The idea is to cancel all the uncertainties $\sigma(s)$ with the help of the control signal $u(s)$. Therefore, the control signal is defined as [10]:

$$u(s) = C(s)\{r(s) - \hat{\sigma}(s)\} \tag{E.21}$$

where $C(s)$ is a first order filter in this application:

$$C(s) = \frac{c}{s+c} \tag{E.22}$$

If Equation (E.21) is substituted in (E.19) then:

$$y(s) = M(s)C(s)r(s) + M(s)\{\sigma(s) - C(s)\hat{\sigma}(s)\} \tag{E.23}$$

If $\sigma(s)$ is perfectly estimated then $\hat{\sigma}(s) = \sigma(s)$ and the disturbances will be rejected only in the bandwidth of $C(s)$.

The adaptive estimate $\hat{\sigma}(t)$ is updated using a projection algorithm that ensures boundedness of the estimate within a given ball [10]:

$$\dot{\hat{\sigma}}(t) = \Gamma \text{Proj}\,(\hat{\sigma}(t), -\tilde{y}(t)) \tag{E.24}$$

where $\tilde{y}(t) = \hat{y}(t) - y(t)$ is the estimation error of the output.

There are 3 parameters to set for the L1 adaptive output feedback controller, i.e. the desired closed loop time constant $1/m$, the adaptation gain Γ and the eigenfrequency of the control filter c.

The reference model and the control signal filter can be designed systematically [125]. Assuming perfect knowledge of disturbances, an ideal system $y(s)$ can be built and used for tuning [125]:

$$y(s) = H(s)C(s)r(s) + H(s)\{1 - C(s)\}d(s) \tag{E.25}$$

where $H(s)C(s)$ is the transfer function from $r(s)$ to $y(s)$ and $H(s)$ is defined as:

$$H(s) = \frac{A(s)M(s)}{C(s)A(s) + \{1 - C(s)\}M(s)} \tag{E.26}$$

Parameters m and c must be chosen such that $H(s)$ is stable and the following L1 norm holds [10]:

$$||G(s)||_{\mathscr{L}_1} L < 1 \tag{E.27}$$

where $G(s) = H(s)\{1 - C(s)\}$ is the transfer function from $d(s)$ to $y(s)$ and L is the Lipschitz constant required to guarantee BIBO stability (Lemma 4.1.1 in [10]).

By considering $A(s)$ as the linearized system around the nominal point, a (m,c) map can be built in order to facilitate the choice of these two parameters. The closed loop system has several poles, which can cause oscillations if they are complex. As a criterion for (m,c) determination, the worst damping ratio ζ should be higher than 0.7 in order to ensure acceptable oscillations. The damping ratio ζ is plotted as a function of (m,c) in the top plot of Figure E.7.

In order to accelerate the system, parameter m should be increased. To reduce oscillations in the system response, the bandwidth of filter $C(s)$ needs to be enlarged

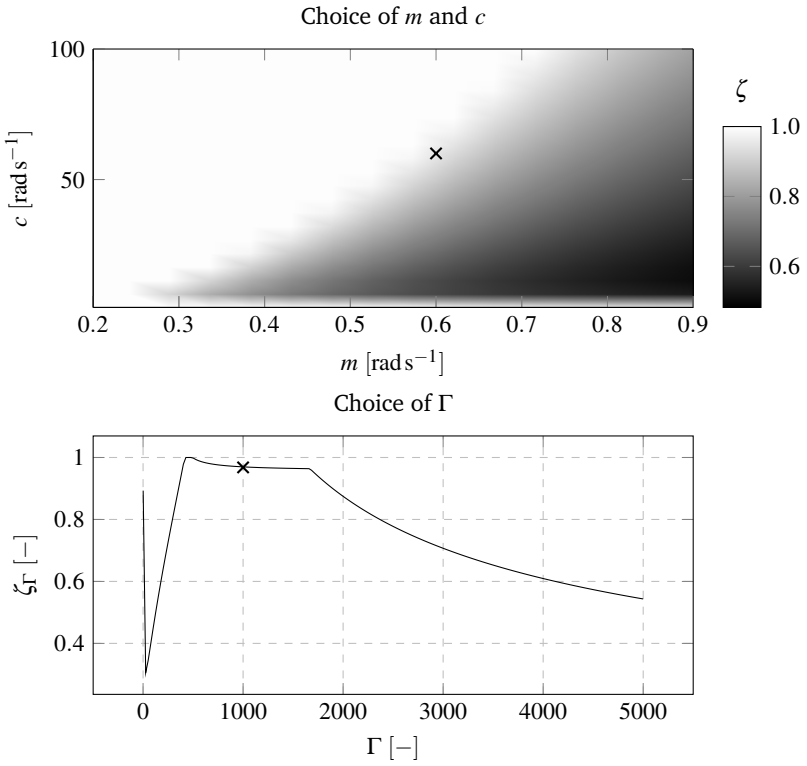


Figure E.7: Tuning of the L1 adaptive controller.

such that a more responsive control action would be allowed. The pair $(0.6, 60)$ is chosen for this application, which corresponds to a $\zeta > 0.9$:

$$m = 0.6 \qquad c = 60 \tag{E.28}$$

The model reference has a time constant that corresponds to 1.6h. The Lipschitz constant, for the selected m and c, is computed to be:

$$L = 78.6851 \tag{E.29}$$

which ensures a BIBO closed loop system according to Lemma 4.1.1 [10].

In theorem 4.1.1 from [10] it is shown that the tracking error between the real and the ideal system, which assumes perfect knowledge of the disturbance, is uniformly bounded with respect to a constant proportional to $1/\sqrt{\Gamma}$. Therefore, a high adaptation gain Γ is desired. At the same time, the stability and dynamics of $\hat{\sigma}$ are dependent on Γ. The transfer function from $r(s)$ and $d(s)$ to $\hat{\sigma}(s)$ is [125]:

$$\hat{\sigma} = F(s)[C(s)\{A(s) - M(s)\}r(s) + A(s)d(s)] \tag{E.30}$$

where $F(s)$ is identified as:

$$F(s) = \frac{1}{\frac{1}{\Gamma}s + C(s)A(s) + \{1 - C(s)\}M(s)} \tag{E.31}$$

Γ should be chosen such that $F(s)$ is stable. To reduce the high frequency noise due to adaptation, Γ can be set such that the worst damping ratio of $F(s)$, denoted as ζ_Γ, would be greater than 0.7. A plot of ζ_Γ as a function of Γ is shown in the bottom plot of Figure E.7. A value $\Gamma = 1000$ would be a good choice for the current application ensuring a damping ratio of $\hat{\sigma}$ close to 1 and a relatively high adaptation gain.

Following the described procedure, the L1 controller can be tuned in different pH operation points and parameters m, c and Γ could be adjusted in real time in order to maximize performances. Table E.4 contains the controller parameters in 3 different operating points.

Table E.4: L1 controller parameters in multiple operational points.

pH	m	c	Γ
3	1	50	800
5	0.6	60	1000
7	4	80	500

The feed-forward term developed in the classical control section is also added to the L1 controller to test its efficiency.

E.6 Benchmark Tests

Reference tracking is tested by performing a square wave of magnitude 2 as in Table E.5. The tests cover the entire pH interval $3-7$, which includes most of the nonlinearity from the titration curve. Even though enzymes normally operate around $pH = 5$, it is of theoretical interest to test the controller for a wider range of operating points.

Table E.5: Reference tracking square wave scenario.

#	r_{pH}	Time interval
1.	5	$0-20$
2.	$\uparrow 7$	$20-40$
3.	$\downarrow 5$	$40-60$
4.	$\downarrow 3$	$60-80$
5.	$\uparrow 5$	$80-100$

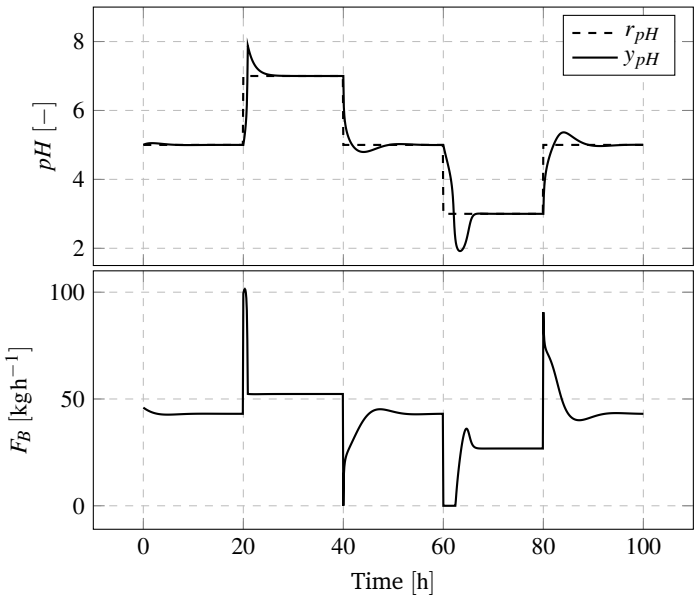


Figure E.8: Reference tracking with a PI controller. Top plot shows the pH measurement and the reference level while the bottom plot displays the control effort or the addition of base.

The disturbance rejection scenario is presented in Table E.6 and includes white noise perturbations with 0 mean and standard deviation σ on the feed concentrations of acetic acid, base and unknown buffers.

Table E.6: White noise disturbance rejection scenarios.

#	Scenario	σ	Time interval
1.	Acid disturbances	$1\,\mathrm{g\,kg^{-1}}$	$0-133$
2.	Base disturbances	$30\,\mathrm{g\,kg^{-1}}$	$133-266$
3.	Unknown buffers	$0.02\,\mathrm{mol\,L^{-1}}$	$266-400$

Several theoretical bell shaped enzymatic activity curves were shown in Figure E.2. The maximum efficiency of the enzymatic process is considered to be 1 and is reached when the pH level equals the optimal value. A small deviation from the optimal level can cause a significant drop in the enzymatic activity. The monitoring cost function is constructed by integrating the deviations from maximum enzymatic activity within a time window:

$$J = \int_{t_0}^{t_1} \{1 - E(y_{pH})\} d\tau \tag{E.32}$$

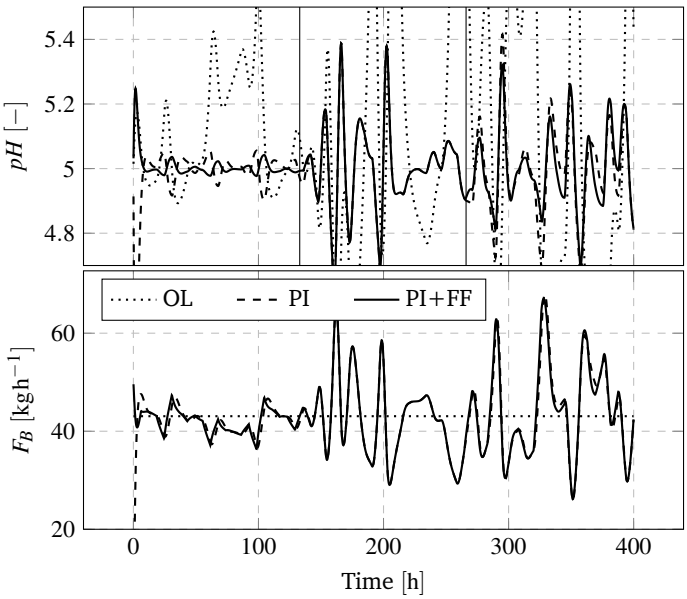


Figure E.9: Disturbance rejection with a PI controller. Top plot shows the reactor pH level in 3 cases: open loop (OL), PI controller (PI) and PI controller with feed-forward action (PI+FF). The bottom plot displays the control effort.

where t_0 is the initial time, t_1 is the final time and $E(y_{pH})$ is the enzymatic activity associated with pH level y_{pH} and can be approximated with a Gaussian bell-shaped curve:

$$E(y_{pH}) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp^{-0.5\left(\frac{y_{pH}-\mu}{\sigma}\right)^2} \tag{E.33}$$

where $\sigma = 0.2$ and $\mu = 5$ for this application. These values correspond to E_2 from Figure E.2. The enzymatic activity is usually experimentally determined and can be represented as a table based map in reality.

Both control strategies are tested in the scenarios described above and the results are commented in the next section.

E.7 Results

The results for the reference tracking scenario can be viewed in Figures E.8 and E.10.

The classical control strategy has overshoots that increase as the system moves further from the design point. On the other hand, the L1 controller has a significant overshoot only in the neutral area of $pH = 7$. Otherwise, the L1 controller responds better than the PI controller following the model reference. The overshoots could

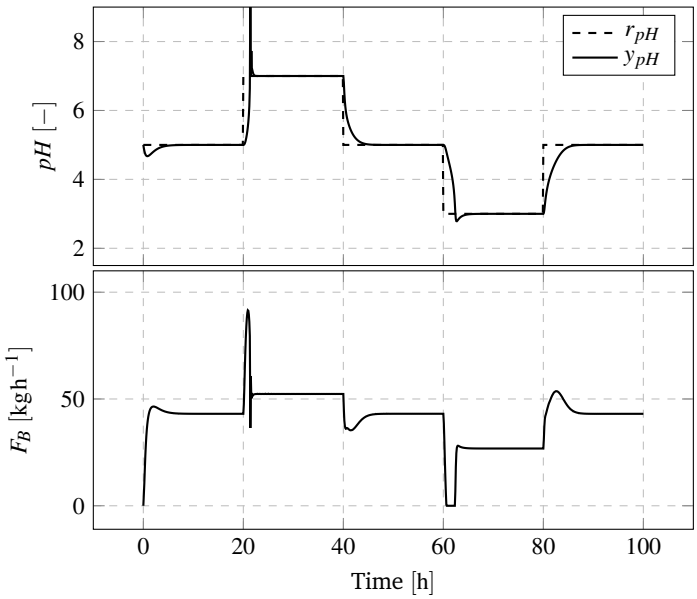


Figure E.10: Reference tracking with an L1 controller. Top plot shows the pH measurement and the reference level while the bottom plot displays the control effort or the addition of base.

be accommodated by a reference filter in both cases. The control signal is within acceptable limits.

The disturbance rejection scenario is shown in Figures E.9 and E.11 and the evaluation of the cost function is performed in Table E.7. In the case of measured disturbances, i.e. acetic acid disturbances, the L1 controller with feed-forward performs better than the PI controller leading to a full rejection of these disturbances. The nonlinear feed-forward term significantly helps both controllers.

In the second scenario, i.e. base disturbances, the L1 controller has better results leading to smaller variations of the pH level, thus to a much lower J value. The feed-forward term does not help in this case.

In the last scenario, i.e. unknown buffers, the L1 controller outperforms the classical PI. Overall, regarding variations in the pH level, the L1 controller has better performances.

E.8 Conclusions

In this article, for the first time to our knowledge, a pH model of the enzymatic liquefaction process has been developed. Reference tracking and disturbance rejec-

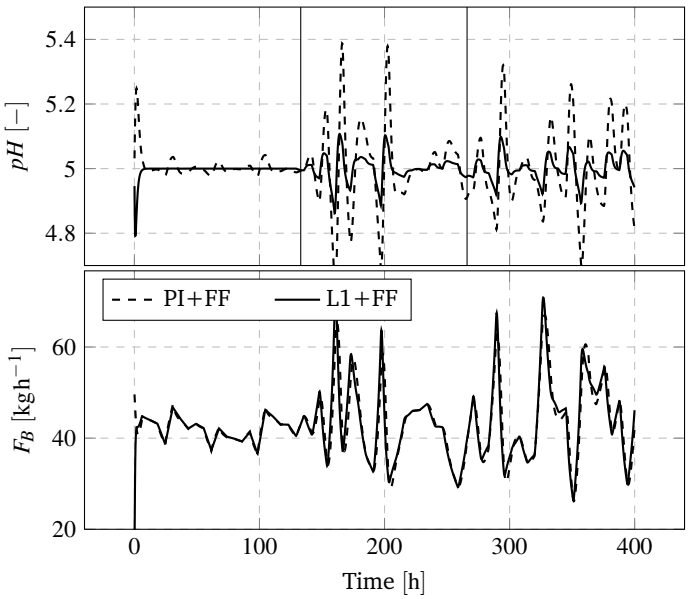


Figure E.11: Disturbance rejection with an L1 adaptive controller. Top plot shows the reactor pH level in 2 cases: PI+FF and L1 controller with feed-forward (L1+FF). The bottom plot displays the control effort.

Table E.7: Performance cost function.

Scenario	PI	PI+FF	L1+FF
1.	62.87	18.67	6.93
2.	188.19	187.16	24.42
3.	237.39	184.95	21.44

tion scenarios were formulated resembling the production mode of a large scale bioethanol plant and both measured and unmeasured disturbances were considered.

As classical control is wide spread within industry, a PI controller with feed-forward action was first designed. A novel L1 adaptive output feedback controller was then built and tuned in a systematic way in order to ensure high performance. The feed-forward term was also added to the L1 controller. In the case of reference tracking, the PI controller performed well with small overshoots that increased as the process moved further from the design point. Re-tuning of the controller is necessary when switching to a different pH level in order to preserve performances. The L1 controller proved to behave similarly regardless of the nominal pH level except in the neutral highly nonlinear area of $pH = 7$. However, the closed loop system remained stable.

In the case of measured disturbances, both control strategies were effective and it was shown that the feed-forward term considerably improves the system response. In the case of unmeasured disturbances, the L1 adaptive controller had better performances.

A new cost function was derived from the enzymatic bell-shaped activity curve to assess closed loop performances. This cost function was further used to monitor the efficiency of the enzymatic process.

Finally, a tuning method was proposed for the L1 controller that proved to be very effective for this application. This tuning procedure is an enhancement of the method presented in [125].

Acknowledgements

The close collaboration and very helpful suggestions and detailed comments received from Dr. Jakob M. Jensen on this research are gratefully acknowledged.

Bibliography

[1] R. Datta, M. A. Maher, C. Jones, and R. W. Brinker. “Ethanol-the primary renewable liquid fuel”. *Journal of Chemical Technology and Biotechnology* 86 (2011), pp. 473–480. DOI: 10.1002/jctb.2580.

[2] R. C. Brown, T. Brown, S. Capareda, B. Dale, D. Edwards, V. Estes, C. Granda, M. Holtzapple, S. Lonkar, and C. E. Wyman. “SBE Supplement: Lignocellulosic Biofuels”. *AIChE CEP* (2015), pp. 33–64.

[3] J. Larsen, M. Ø. Haven, and L. Thirup. “Inbicon makes lignocellulosic ethanol a commercial reality”. *Biomass and Bioenergy* 46 (Nov. 2012), pp. 36–45. DOI: 10.1016/j.biombioe.2012.03.033.

[4] J. Larsen, M. Ø. Petersen, L. Thirup, H. W. Li, and F. K. Iversen. “The IBUS Process - Lignocellulosic Bioethanol Close to a Commercial Reality”. *Chemical Engineering & Technology* 31 (May 2008), pp. 765–772. DOI: 10.1002/ceat.200800048.

[5] R. M. Prunescu, M. Blanke, J. G. Jakobsen, and G. Sin. “Dynamic modeling and validation of a biomass hydrothermal pretreatment process - A demonstration scale study”. *AIChE Journal* (2015). DOI: 10.1002/aic.14954.

[6] R. M. Prunescu and G. Sin. “Dynamic modeling and validation of a lignocellulosic enzymatic hydrolysis process - A demonstration scale study”. *Bioresource Technology* 150 (Dec. 2013), pp. 393–403. DOI: 10.1016/j.biortech.2013.10.029.

[7] R. M. Prunescu, M. Blanke, J. G. Jakobsen, and G. Sin. “Model-Based Plantwide Optimization of a Large Scale Lignocellulosic Bioethanol Plant”. *Submitted to AIChE Journal* (2015).

[8] G. Sin, A. S. Meyer, and K. V. Gernaey. “Assessing reliability of cellulose hydrolysis models to support biofuel process design - Identifiability and uncertainty analysis”. *Computers & Chemical Engineering* 34 (Sept. 2010), pp. 1385–1392. DOI: 10.1016/j.compchemeng.2010.02.012.

[9] R. Morales-Rodriguez, A. S. Meyer, K. V. Gernaey, and G. Sin. "A framework for model-based optimization of bioprocesses under uncertainty: Lignocellulosic ethanol production case". *Computers & Chemical Engineering* 42 (July 2012), pp. 115–129. DOI: 10.1016/j.compchemeng.2011.12.004.

[10] N. Hovakimyan and C. Cao. *L1 Adaptive Control Theory*. SIAM, 2010.

[11] R. M. Prunescu, M. Blanke, and G. Sin. "Modelling and L1 Adaptive Control of pH in Bioethanol Enzymatic Process". *Proceedings of the 2013 American Control Conference*. Washington D.C., USA, 2013, pp. 1888–1895.

[12] R. M. Prunescu, M. Blanke, and G. Sin. "Modelling and L1 Adaptive Control of Temperature in Biomass Pretreatment". *Proceedings of the 52nd IEEE Conference on Decision and Control*. Florence, Italy, 2013, pp. 3152–3159.

[13] J. Conti, P. Holtberg, J. A. Beamon, S. Napolitano, M. A. Schaal, J. T. Turnure, and L. Westfall. *International Energy Outlook 2013*. Tech. rep. 2013. DOI: EIA-0484(2013).

[14] R. Guseo, A. Dalla Valle, and M. Guidolin. "World Oil Depletion Models: Price effects compared with strategic or technological interventions". *Technological Forecasting and Social Change* 74.4 (2007), pp. 452–469. DOI: 10.1016/j.techfore.2006.01.004.

[15] K. Mohaddes. "Econometric Modelling of World Oil Supplies: Terminal Price and the Time to Depletion". *OPEC Energy Review* (2012), pp. 162–193.

[16] C. McGlade and P. Ekins. "The geographical distribution of fossil fuels unused when limiting global warming to 2C". *Nature* 517.7533 (2014), pp. 187–190. DOI: 10.1038/nature14016.

[17] T. Gasser, C. Guivarch, K. Tachiiri, C. D. Jones, and P. Ciais. "Negative emissions physically needed to keep global warming below 2C". *Nature Communications* 6 (2015), p. 7958. DOI: 10.1038/ncomms8958.

[18] N. Armaroli and V. Balzani. "The legacy of fossil fuels". *Chemistry - An Asian Journal* 6.3 (2011), pp. 768–784. DOI: 10.1002/asia.201000797.

[19] United Nations. *Kyoto Protocol To the United Nations Framework*. 1998. DOI: 10.1111/1467-9388.00150.

[20] M. Wang, M. Wu, and H. Huo. "Life-cycle energy and greenhouse gas emission impacts of different corn ethanol plant types". *Environmental Research Letters* 2.2 (2007), pp. 1–13. DOI: 10.1088/1748-9326/2/2/024001.

[21] M. A. Martin. "First generation biofuels compete". *New Biotechnology* 27.5 (2010), pp. 596–608. DOI: 10.1016/j.nbt.2010.06.010.

[22] Novozymes A/S. “Beta Renewables opens biofuels plant in Italy”. *Focus on Catalysts* (2013), p. 6. DOI: 10.1016/S1351-4180(13)70461-6.

[23] G. Dragone, B. Fernandes, A. Vicente, and J. Teixeira. “Third generation biofuels from microalgae”. *Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology* (2010), pp. 1355–1366.

[24] F. Alam, S. Mobin, and H. Chowdhury. “Third Generation Biofuel from Algae”. *Procedia Engineering* 105 (2015), pp. 763–768. DOI: 10.1016/j.proeng.2015.05.068.

[25] J. Miranowski and A. Rosburg. “Long-term biofuel projections under different oil price scenarios”. *AgBioForum* 16.1 (2013), pp. 79–87.

[26] P. Cheali, J. A. Posada, K. V. Gernaey, and G. Sin. “Upgrading of lignocellulosic biorefinery to value-added chemicals: Sustainability and economics of bioethanol-derivatives”. *Biomass and Bioenergy* 75 (2015), pp. 282–300. DOI: 10.1016/j.biombioe.2015.02.030.

[27] C. E. Wyman and A. J. Ragauskas. “Lignin Bioproducts to Enable Biofuels”. *Biofuels, Bioproducts and Biorefining* 9.5 (2015), pp. 447–449. DOI: 10.1002/bbb.1582.

[28] A. J. Ragauskas, G. T. Beckham, M. J. Biddy, R. Chandra, F. Chen, M. F. Davis, B. H. Davison, R. a. Dixon, P. Gilna, M. Keller, P. Langan, A. K. Naskar, J. N. Saddler, T. J. Tschaplinski, G. a. Tuskan, and C. E. Wyman. “Lignin valorization: improving lignin processing in the biorefinery.” *Science (New York, N.Y.)* 344.6185 (2014), p. 1246843. DOI: 10.1126/science.1246843.

[29] Innovation Fund Denmark. *Innovation Fund Denmark 2015 Strategy*. Tech. rep. 2015.

[30] D. A. Mitchell, N. Krieger, and M. Berovič. *Solid-State Fermentation Bioreactors*. Springer, 2006. DOI: 10.1007/3540312854.

[31] N. Silver. *The Signal and the Noise: The Art and Science of Prediction*. Allen Lane, 2012.

[32] R. M. Prunescu, M. Blanke, J. G. Jakobsen, and G. Sin. “Dynamic Modeling, Advanced Control, Diagnosis and Optimization of Large-Scale Lignocellulosic Biorefineries”. *Proceedings of the AIChE 2015 Annual Meeting*. Salt Lake City, UT, USA, 2015.

[33] R. M. Prunescu, M. Blanke, J. G. Jakobsen, and G. Sin. “Plantwide Model-Based Optimization of a Large Scale Second Generation Biorefinery”. *Proceedings of the AIChE 2015 Annual Meeting*. Salt Lake City, UT, USA, 2015.

[34] R. M. Prunescu, M. Blanke, J. G. Jakobsen, and G. Sin. “Model-Based Filtering of Large-Scale Datasets - A Biorefinery Application”. *Proceedings of the AIChE 2014 Annual Meeting*. Atlanta, GA, USA, 2014.

[35] R. M. Prunescu and G. Sin. “Dynamic Simulation, Sensitivity and Uncertainty Analysis of a Demonstration Scale Lignocellulosic Enzymatic Hydrolysis Process”. *Proceedings of the AIChE 2014 Annual Meeting*. Atlanta, GA, USA, 2014.

[36] R. M. Prunescu, M. Blanke, and G. Sin. “Advances in Monitoring, Diagnosis and Control of Biorefineries”. *Proceedings of the 9th World Congress of Chemical Engineering*. Seoul, South Korea, 2013.

[37] A. Sluiter, B. Hames, R. Ruiz, C. Scarlata, J. Sluiter, D. Templeton, and D. Crocker. “Determination of structural carbohydrates and lignin in biomass”. *Technical Report NREL/TP-510-42618* (2008). DOI: NREL/TP-510-42618.

[38] C. Carrasco, D. Cuno, K. Carlqvist, M. Galbe, and G. Lidén. “SO2-catalysed steam pretreatment of quinoa stalks”. *Journal of Chemical Technology & Biotechnology* 90.1 (2015), pp. 64–71. DOI: 10.1002/jctb.4286.

[39] S. Collins, N. Wellner, I. Martinez Bordonado, A. L. Harper, C. N. Miller, I. Bancroft, and K. W. Waldron. “Variation in the chemical composition of wheat straw: the role of tissue ratio and composition”. *Biotechnology for Biofuels* 7.1 (2014), p. 121. DOI: 10.1186/PREACCEPT-4657379161257175.

[40] J. B. Kristensen, L. G. Thygesen, C. Felby, H. Jørgensen, and T. Elder. “Cell-wall structural changes in wheat straw pretreated for bioethanol production.” *Biotechnology for biofuels* 1 (Jan. 2008), pp. 1–9. DOI: 10.1186/1754-6834-1-5.

[41] M. A. T. Hansen and J. B. Kristensen. “Pretreatment and enzymatic hydrolysis of wheat straw (Triticum aestivum L.)–The impact of lignin relocation and plant tissues on enzymatic accessibility”. *Bioresource technology* 102 (Feb. 2011), pp. 2804–2811. DOI: 10.1016/j.biortech.2010.10.030.

[42] N. D. Weiss, J. D. Farmer, and D. J. Schell. “Impact of corn stover composition on hemicellulose conversion during dilute acid pretreatment and enzymatic cellulose digestibility of the pretreated solids.” *Bioresource technology* 101 (Jan. 2010), pp. 674–678. DOI: 10.1016/j.biortech.2009.08.082.

[43] D. W. Templeton, C. J. Scarlata, J. B. Sluiter, and E. J. Wolfrum. “Compositional analysis of lignocellulosic feedstocks. 2. Method uncertainties.” *Journal of agricultural and food chemistry* 58 (Aug. 2010), pp. 9054–9062. DOI: 10.1021/jf100807b.

[44] D. Chiaramonti, M. Prussi, S. Ferrero, L. Oriani, P. Ottonello, P. Torre, and F. Cherchi. “Review of pretreatment processes for lignocellulosic ethanol production, and development of an innovative method”. *Biomass and Bioenergy* 46 (Nov. 2012), pp. 25–35. DOI: 10.1016/j.biombioe.2012.04.020.

[45] J. Larsen and N. Henriksen. *Status for the Inbicon technology by end of 2014*. Tech. rep. 2014, pp. 1–8.

[46] B. Yang and C. E. Wyman. “Pretreatment: The key to unlocking low-cost cellulosic ethanol”. *Biofuels, Bioproducts and Biorefining* 2.1 (2008), pp. 26–40. DOI: 10.1002/bbb.49.

[47] C. Carrasco, H. Baudel, M. Peñarrieta, C. Solano, L. Tejeda, C. Roslander, M. Galbe, and G. Lidén. “Steam pretreatment and fermentation of the straw material "Paja Brava" using simultaneous saccharification and co-fermentation”. *Journal of Bioscience and Bioengineering* 111.2 (2011), pp. 167–174. DOI: 10.1016/j.jbiosc.2010.10.009.

[48] I. Cybulska, G. Brudecki, J. E. Schmidt, and M. H. Tomsen. “Organosolv Fractionation of Palm Tree Residues”. *Energy Procedia* 75 (2015), pp. 742–747. DOI: 10.1016/j.egypro.2015.07.503.

[49] I. Cybulska, G. Brudecki, and H. Lei. “Hydrothermal Pretreatment of Lignocellulosic Biomass”. *Green Biomass Pretreatment for Biofuels Production*. Springer Netherlands, 2013. Chap. 4, pp. 87–106. DOI: 10.1007/978-94-007-6052-3_4.

[50] M. Ø. Petersen, J. Larsen, and M. H. Thomsen. “Optimization of hydrothermal pretreatment of wheat straw for production of bioethanol at low water consumption without addition of chemicals”. *Biomass and Bioenergy* 33 (May 2009), pp. 834–840. DOI: 10.1016/j.biombioe.2009.01.004.

[51] Q. Qing, B. Yang, and C. E. Wyman. “Xylooligomers are strong inhibitors of cellulose hydrolysis by enzymes.” *Bioresource technology* 101 (Dec. 2010), pp. 9624–9630. DOI: 10.1016/j.biortech.2010.06.137.

[52] M. Cantarella, L. Cantarella, A. Gallifuoco, A. Spera, and F. Alfani. “Effect of inhibitors released during steam-explosion treatment of poplar wood on subsequent enzymatic hydrolysis and SSF”. *Biotechnology progress* 20 (2004), pp. 200–206. DOI: 10.1021/bp0257978.

[53] P. Sannigrahi, D. H. Kim, S. Jung, and A. Ragauskas. “Pseudo-lignin and pretreatment chemistry”. *Energy & Environmental Science* 4 (2011), pp. 1306–1310. DOI: 10.1039/c0ee00378f.

[54] C. Felby, J. Larsen, H. Jørgensen, and J. Vibe-Pedersen. *Enzymatic Hydrolysis of Biomasses Having High Dry Matter (DM) Content*. 2006.

[55] H. Jørgensen, J. Vibe-Pedersen, J. Larsen, and C. Felby. "Liquefaction of lignocellulose at high solids concentrations". *Biotechnology and Bioengineering* 96 (2007), pp. 862–870. DOI: 10.1002/bit.

[56] B. Palmqvist and G. Lidén. "Torque measurements reveal large process differences between materials during high solid enzymatic hydrolysis of pretreated lignocellulose". *Biotechnology for Biofuels* 5.1 (2012), p. 57. DOI: 10.1186/1754-6834-5-57.

[57] Novozymes A/S. "Cellic® CTec2 and HTec2 - Enzymes for hydrolysis of lignocellulosic materials" (2010), pp. 1–9.

[58] Novozymes A/S. *Cellulosic ethanol Novozymes Cellic CTec3 - Secure your plant's lowest total cost*. Tech. rep. 2012, pp. 1–6.

[59] Novozymes A/S. *Cellulosic ethanol - Novozymes Cellic HTec3*. Tech. rep. 2012, pp. 1–3.

[60] K. L. Kadam, E. C. Rydholm, and J. D. McMillan. "Development and validation of a kinetic model for enzymatic saccharification of lignocellulosic biomass." *Biotechnology progress* 20 (2004), pp. 698–705. DOI: 10.1021/bp034316x.

[61] Y. Zhang, J. L. Xu, H. J. Xu, Z. H. Yuan, and Y. Guo. "Cellulase deactivation based kinetic modeling of enzymatic hydrolysis of steam-exploded wheat straw". *Bioresource Technology* 101.21 (2010), pp. 8261–8266. DOI: 10.1016/j.biortech.2010.06.015.

[62] L. Ljung. *System Identification: Theory for the User*. 2nd Editio. Prentice Hall, 1999.

[63] E. Walter and L. Pronzato. *Identification of parametric models from experimental data*. Springer, 1997.

[64] R. Brun, P. Reichert, and H. H. R. Künsch. "Practical identifiability analysis of large environmental simulation models". *Water Resources Research* 37 (2001), pp. 1015–1030. DOI: 10.1029/2000WR900350.

[65] J. Helton and F. Davis. "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems". *Reliability Engineering & System Safety* 81 (July 2003), pp. 23–69. DOI: 10.1016/S0951-8320(03)00058-9.

[66] G. Sin, K. V. Gernaey, M. B. Neumann, M. C. M. van Loosdrecht, and W. Gujer. “Global sensitivity analysis in wastewater treatment plant model applications: Prioritizing sources of uncertainty”. *Water Research* 45 (Jan. 2011), pp. 639–651. DOI: 10.1016/j.watres.2010.08.025.

[67] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis. The Primer*. Chichester, UK: John Wiley & Sons, Ltd, 2007. DOI: 10.1002/9780470725184.

[68] J. Price, M. Nordblad, J. M. Woodley, and J. K. Huusom. “Application of Uncertainty and Sensitivity Analysis to a Kinetic Model for Enzymatic Biodiesel Production”. *12th IFAC Symposium on Computer Applications in Biotechnology*. Mumbai, 2013, pp. 149–156. DOI: 10.3182/20131216-3-IN-2044.00060.

[69] J. R. Cooper and R. B. Dooley. *Revised release on the IAPWS industrial formulation 1997 for the thermodynamic properties of water and steam*. The International Association for the Properties of Water and Steam, 2007.

[70] R. B. Bird, W. E. Stewart, and E. N. Lightfoot. *Transport phenomena*. John Wiley & Sons, 2007.

[71] R. M. Prunescu, M. Blanke, J. M. Jensen, and G. Sin. “Temperature Modelling of the Biomass Pretreatment Process”. *Proceedings of the 17th Nordic Process Control Workshop*. Copenhagen, Denmark, 2012, pp. 8–17. DOI: 10.13140/2.1.1512.1287.

[72] O. Egeland and J. T. Gravdahl. *Modeling and simulation for automatic control*. Marine Cybernetics, 2002.

[73] B. Lavarack, G. Griffin, and D. Rodman. “The acid hydrolysis of sugarcane bagasse hemicellulose to produce xylose, arabinose, glucose and other products”. *Biomass and Bioenergy* 23 (2002), pp. 367–380.

[74] M. Mauricio-Iglesias, J. K. Huusom, and K. Gernaey. “State Estimation in Fermentation of Lignocellulosic Ethanol. Focus on the Use of pH Measurements”. *Proceedings of the 25th European Symposium on Computer Aided Process Engineering*. 2015, pp. 1769–1774.

[75] R. Luff, M. Haeckel, and K. Wallmann. “Robust and fast FORTRAN and MATLAB libraries to calculate pH distributions in marine systems”. *Computers and Geosciences* 27 (2001), pp. 157–169. DOI: 10.1016/S0098-3004(00)00097-2.

[76] J. Price, M. Nordblad, J. M. Woodley, and J. K. Huusom. “Real-time model based process monitoring of enzymatic biodiesel production”. *Biotechnology Progress* 31.2 (2015), pp. 585–595. DOI: 10.1002/btpr.2030.

[77] M. Michelin, E. Ximenes, M. de Lourdes Teixeira de Moraes Polizeli, and M. R. Ladisch. “Effect of phenolic compounds from pretreated sugarcane bagasse on cellulolytic and hemicellulolytic activities”. *Bioresource Technology* (2015), pp. 8–11. DOI: 10.1016/j.biortech.2015.08.120.

[78] K. K. Podkaminer, X. Shao, D. a. Hogsett, and L. R. Lynd. “Enzyme inactivation by ethanol and development of a kinetic model for thermophilic simultaneous saccharification and fermentation at 50C with Thermoanaerobacterium saccharolyticum ALK2”. *Biotechnology and Bioengineering* 108.6 (2011), pp. 1268–1278. DOI: 10.1002/bit.23050.

[79] S. Mutturi and G. Lidén. “Model-based estimation of optimal temperature profile during simultaneous saccharification and fermentation of Arundo donax”. *Biotechnology and Bioengineering* 111.5 (2014), pp. 866–875. DOI: 10.1002/bit.25165.

[80] K. Olofsson, A. Rudolf, and G. Lidén. “Designing simultaneous saccharification and fermentation for improved xylose conversion by a recombinant strain of Saccharomyces cerevisiae”. *Journal of Biotechnology* 134.1-2 (2008), pp. 112–120. DOI: 10.1016/j.jbiotec.2008.01.004.

[81] J. J. Downs and S. Skogestad. “An industrial and academic perspective on plantwide control”. *Annual Reviews in Control* 35 (2011), pp. 99–110. DOI: 10.1016/j.arcontrol.2011.03.006.

[82] D. E. Seborg, D. A. Mellichamp, T. F. Edgar, and F. J. Doyle. “Process Dynamics and Control”. *Process Dynamics and Control* (2010), p. 2010.

[83] M. Ø. Haven, J. Lindedam, M. D. Jeppesen, M. Elleskov, A. C. Rodrigues, M. Gama, H. Jørgensen, and C. Felby. “Continuous recycling of enzymes during production of lignocellulosic bioethanol in demonstration scale”. *Applied Energy* 159 (2015), pp. 188–195. DOI: 10.1016/j.apenergy.2015.08.062.

[84] V. Gera, M. Panahi, S. Skogestad, and N. Kaistha. “Economic Plantwide Control of the Cumene Process”. *Industrial & Engineering Chemistry Research* 52.2 (2013), pp. 830–846. DOI: 10.1021/ie301386h.

[85] S. Mutturi and G. Lidén. “Effect of Temperature on Simultaneous Sacchari fi cation and Fermentation of Pretreated Spruce and Arundo”. *Industrial & Engineering Chemistry Research* 52 (2013), pp. 1244–1251. DOI: dx.doi.org/10.1021/ie302851w.

[86] O. Johnsson, J. Andersson, G. Lidén, C. Johnsson, and T. Hägglund. “Feed rate control in fed-batch fermentations based on frequency content analysis”. *Biotechnology Progress* 29.3 (2013), pp. 817–824. DOI: 10.1002/btpr.1727.

[87] J. K. Huusom. “Challenges and opportunities in integration of design and control”. *Computers & Chemical Engineering* 81 (2015), pp. 138–146. DOI: 10.1016/j.compchemeng.2015.03.019.

[88] C. Cao and N. Hovakimyan. “Design and Analysis of a Novel Adaptive Control Architecture With Guaranteed Transient Performance”. *Automatic Control IEEE Transactions on* 53.2 (2008), pp. 586–591. DOI: 10.1109/TAC.2007.914282.

[89] M. Mauricio-Iglesias, J. Gottschalck-Andersen, and K. Gernaey. “Modelling and optimization of C5 and C6 fermentation: focus on pH impact and inhibitors effect”. *10th European Symposium on Biochemical Engineering Sciences and 6th International Forum on Industrial Bioprocesses*. Lille, France, 2014.

[90] S. Skogestad. “Simple analytic rules for model reduction and PID controller tuning”. *Modeling, Identification and Control* 25.2 (2004), pp. 85–120.

[91] B. C. Chachuat. *Nonlinear and Dynamic Optimization: From Theory to Practice*. 2009.

[92] F. Logist, M. Vallerio, B. Houska, M. Diehl, and J. Van Impe. “Multi-objective optimal control of chemical processes using ACADO toolkit”. *Computers & Chemical Engineering* 37 (2012), pp. 191–199. DOI: 10.1016/j.compchemeng.2011.11.002.

[93] F. Logist, D. Telen, B. Houska, M. Diehl, and J. Van Impe. “Multi-objective optimal control of dynamic bioprocesses using ACADO Toolkit”. *Bioprocess and Biosystems Engineering* 36.2 (2013), pp. 151–164. DOI: 10.1007/s00449-012-0770-9.

[94] D. Bonvin and B. Srinivasan. “On the role of the necessary conditions of optimality in structuring dynamic real-time optimization schemes”. *Computers & Chemical Engineering* 51 (2013), pp. 172–180. DOI: 10.1016/j.compchemeng.2012.07.012.

[95] G. François, B. Srinivasan, and D. Bonvin. “Comparison of six implicit real-time optimization schemes”. *Journal Europeen des Systemes Automatises* 46.2-3 (2012), pp. 291–305. DOI: 10.3166/JESA.46.291-305.

[96] R. Morales-Rodriguez, A. S. Meyer, K. V. Gernaey, and G. Sin. "Dynamic model-based evaluation of process configurations for integrated operation of hydrolysis and co-fermentation for bioethanol production from lignocellulose." *Bioresource technology* 102 (Jan. 2011), pp. 1174–1184. DOI: 10.1016/j.biortech.2010.09.045.

[97] R. P. Overend, E. Chornet, and J. A. Gascoigne. "Fractionation of lignocellulosics by steam-aqueous pretreatments". *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 321 (1987), pp. 523–536.

[98] A. Holmberg. "On the practical identifiability of microbial growth models incorporating Michaelis-Menten type nonlinearities". *Mathematical Biosciences* 62 (Nov. 1982), pp. 23–43. DOI: 10.1016/0025-5564(82)90061-X.

[99] M. Power. "The predictive validation of ecological and environmental models". *Ecological Modelling* 68 (July 1993), pp. 33–50. DOI: 10.1016/0304-3800(93)90106-3.

[100] L. Ljung. "Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems". *IEEE Transactions on Automatic Control* 24 (1979), pp. 36–50. DOI: 10.1109/TAC.1979.1101943.

[101] W. W. Zhou and M. Blanke. "Identification of a class of non-linear state space models using rpe techniques". *IEEE Transactions on Automatic Control* 34 (1989), pp. 312–316.

[102] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.

[103] R. G. Brown and P. Y. C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering*. 3rd ed. John Wiley & Sons, 1996.

[104] Y.-H. P. Zhang and L. R. Lynd. "Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems". *Biotechnology and Bioengineering* 88.7 (2004), pp. 797–824. DOI: 10.1002/bit.20282.

[105] D. B. Hodge, M. N. Karim, D. J. Schell, and J. D. McMillan. "Model-based fed-batch for high-solids enzymatic cellulose hydrolysis". *Applied Biochemistry and Biotechnology* 152.1 (2009), pp. 88–107. DOI: 10.1007/s12010-008-8217-0.

[106] S. A. Hosseini and N. Shah. "Enzymatic hydrolysis of cellulose part II: Population balance modelling of hydrolysis by exoglucanase and universal kinetic model". *Biomass and Bioenergy* 35.9 (2011), pp. 3830–3840. DOI: 10.1016/j.biombioe.2011.04.029.

[107] S. A. Hosseini and N. Shah. “Enzymatic hydrolysis of cellulose part II: Population balance modelling of hydrolysis by exoglucanase and universal kinetic model”. *Biomass and Bioenergy* 35.9 (2011), pp. 3830–3840. DOI: 10.1016/j.biombioe.2011.04.029.

[108] S. Naik, V. V. Goud, P. K. Rout, and A. K. Dalai. “Production of first and second generation biofuels: A comprehensive review”. *Renewable and Sustainable Energy Reviews* 14.2 (2010), pp. 578–597. DOI: 10.1016/j.rser.2009.10.003.

[109] D. G. Thomas. “Transport characteristics of suspension: VIII. A note on the viscosity of Newtonian suspensions of uniform spherical particles”. *Journal of Colloid Science* 20 (1965), pp. 267–277.

[110] A. Converti, M. Zilli, and S. Arni. “Estimation of viscosity of highly viscous fermentation media containing one or more solutes”. *Biochemical engineering Journal* 4 (1999), pp. 18–22.

[111] S. Glasstone, K. Laidler, and H. Eyring. *The Theory of Rate Processes: The Kinetics of Chemical Reactions, Viscosity, Diffusion and Electrochemical Phenomena*. McGraw Hill, 1941.

[112] T. McAvoy, E. Hsu, and S. Lowenthal. “Dynamics of pH in controlled stirred tank reactor”. *Industrial and Engineering Chemistry Process Design and Development* 11.1 (1972), pp. 68–70. DOI: 10.1021/i260041a013.

[113] G. Sin and P. a. Vanrolleghem. “Extensions to modeling aerobic carbon degradation using combined respirometric-titrimetric measurements in view of activated sludge model calibration”. *Water Research* 41 (Aug. 2007), pp. 3345–3358. DOI: 10.1016/j.watres.2007.03.029.

[114] A. Albert and E. Serjeant. *Ionization constants of acids and bases. A laboratory manual*. Methuen, 1962.

[115] M. Khan. “Hydrolysis of hemicellulose by commercial enzyme mixtures” (2010), pp. 1–28. DOI: 1402-1552-ISRN:LTU-DUPP--10/040--SE.

[116] I. Ballesteros, M. J. Negro, J. M. Oliva, A. Cabañas, P. Manzanares, and M. Ballesteros. “Ethanol production from steam-explosion pretreated wheat straw.” *Applied biochemistry and biotechnology* 129-132 (2006), pp. 496–508. DOI: 10.1385/ABAB:130:1:496.

[117] C. K. Nitsos, K. A. Matis, and K. S. Triantafyllidis. “Optimization of hydrothermal pretreatment of lignocellulosic biomass in the bioethanol production process”. *ChemSusChem* 6 (Jan. 2013), pp. 110–122. DOI: 10.1002/cssc.201200546.

[118] E. Castro, I. U. Nieves, M. T. Mullinnix, W. J. Sagues, R. W. Hoffman, M. T. Fernández-Sandoval, Z. Tian, D. L. Rockwood, B. Tamang, and L. O. Ingram. "Optimization of dilute-phosphoric-acid steam pretreatment of Eucalyptus benthamii for biofuel production". *Applied Energy* 125 (2014), pp. 76–83. DOI: 10.1016/j.apenergy.2014.03.047.

[119] J. Nielsen, J. Villadsen, and G. Lidén. *Bioreaction Engineering Principles*. Second Edi. Springer, 2003.

[120] M. S. Krishnan, N. W. Ho, and G. T. Tsao. "Fermentation kinetics of ethanol production from glucose and xylose by recombinant Saccharomyces 1400 (pLNH33)". *Applied Biochemistry and Biotechnology* 77-79 (1999), pp. 373–388. DOI: 10.1385/ABAB:78:1-3:373.

[121] R. Alinia, S. Zabihi, F. Esmaeilzadeh, and J. F. Kalajahi. "Pretreatment of wheat straw by supercritical CO2 and its enzymatic hydrolysis for sugar production". *Biosystems Engineering* 107.1 (2010), pp. 61–66. DOI: 10.1016/j.biosystemseng.2010.07.002.

[122] N. Mosier, R. Hendrickson, N. Ho, M. Sedlak, and M. R. Ladisch. "Optimization of pH controlled liquid hot water pretreatment of corn stover". *Bioresource Technology* 96.18 SPEC. ISS. (2005), pp. 1986–1993. DOI: 10.1016/j.biortech.2005.01.013.

[123] P. D. Fullbrook. "Practical limits and prospects (kinetics)". *Industrial Enzymology 2nd Edition*. 1996, pp. 505–540.

[124] C. Cao and N. Hovakimyan. "L1 Adaptive Output Feedback Controller for Systems of Unknown Dimension". *Automatic Control, IEEE Transactions* 53.3 (2008), pp. 815–821. DOI: 10.1109/TAC.2008.919550.

[125] R. Hindman, C. Cao, and N. Hovakimyan. "Designing a high performance, stable L1 adaptive output feedback controller". *Proceedings of the AIAA Guidance, Navigation and Control Conference and Exhibit*. 2007, p. 6644.

[126] A. Limayem and S. C. Ricke. "Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects". *Progress in Energy and Combustion Science* 38.4 (2012), pp. 449–467. DOI: 10.1016/j.pecs.2012.03.002.

[127] J.-P. Ylén. "Measuring, Modeling and Controling the pH Value and the Dynamic Chemical State". PhD thesis. Helsinki University of Technology, 2001.

[128] F. Wan, H. Shang, and L. X. Wang. “Adaptive Fuzzy Control of a pH Process”. *2006 IEEE International Conference on Fuzzy Systems*. 2006, pp. 2377–2384. DOI: 10.1109/FUZZY.2006.1682031.

[129] T. K. Gustafsson and K. V. Waller. “Nonlinear and Adaptive Control of pH”. *Ind. Eng. Chem. Res.* 31.12 (1992), pp. 2681–2693. DOI: 10.1021/ie00012a009.

[130] S.-S. Yoon, T.-W. Yoon, D. R. Yang, and T.-S. Kang. “Indirect adaptive nonlinear control of a pH process”. *Computers & Chemical Engineering* 26.9 (2002), pp. 1223–1230. DOI: 10.1016/S0098-1354(02)00088-1.

[131] J. C. Gomez, A. Jutan, and E. Baeyens. “Wiener model identification and predictive control of a pH neutralisation process”. *Control Theory and Applications, IEE Proceedings* 151.3 (2004), pp. 329–338. DOI: 10.1049/ip-cta:20040438(410)151.

[132] A. Altınten. “Generalized predictive control applied to a pH neutralization process”. *Computers & Chemical Engineering* 31.10 (2007), pp. 1199–1204. DOI: 10.1016/j.compchemeng.2006.10.005.

[133] Z. Fan, C. South, K. Lyford, J. Munsie, P. Van Walsum, and L. R. Lynd. “Conversion of paper sludge to ethanol in a semicontinuous solids-fed reactor”. *Bioprocess and Biosystems Engineering* 26.2 (2003), pp. 93–101. DOI: 10.1007/s00449-003-0337-x.

Technical University of Denmark
Automation and Control (AUT)
Elektrovej Building 326
DK-2800, Kgs. Lyngby
Denmark
Phone: (+45) 45 25 35 76
Email: info@elektro.dtu.dk
www.elektro.dtu.dk

ISBN: N/A